



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 124162

TO: Rita Mitra
Location: Rem 3 B 65
Art Unit: 1653
Tuesday, June 08, 2004

3C70
236

Case Serial Number: 10/023529

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-B55
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Mitra,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
Remsen 1-A-62
Ext. 22524

No art



Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	546	100.0		546	4	Q86YW3	Q86YW3 homo sapien
2	529	96.9		546	4	Q86T54	Q86T54 homo sapien
3	433	79.3		546	4	Q86T86	Q86T86 homo sapien
4	243	44.5		546	4	Q86T85	Q86T85 homo sapien
5	116	21.2		116	4	Q86T86	Q86Y96 homo sapien
6	21	3.8		150	11	Q8C5K1	Q8C5K1 mus musculus
7	19	3.5		241	11	Q8BWJ7	Q8BWJ7 mus musculus
8	19	3.5		241	11	Q8BP11	Q8BP11 mus musculus
9	19	3.5		463	11	Q8BUK2	Q8BUK2 mus musculus
10	19	3.5		524	11	Q8BHN1	Q8BHN1 mus musculus
11	19	3.5		528	4	Q8NUQ3	Q8NUQ3 homo sapien
12	19	3.5		684	11	Q8VBTL	Q8VBTL mus musculus
13	17	3.1		186	4	Q9P0X1	Q9P0X1 homo sapien
14	17	3.1		505	4	Q8N3S2	Q8N3S2 homo sapien
15	17	3.1		676	13	Q919G9	Q919G9 gallus gall
16	17	3.1		715	4	Q8N3L3	Q8N3L3 homo sapien

```

QY 181 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240
DB 181 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240
QY 241 LESLRELQRHNSLKEGVQARBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ 300
DB 241 LESLRELQRHNSLKEGVQARBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ 300
QY 301 ENMELAEERLKKLIEQYELREEHIDKVFPHKDLQOOLVDKLOQQAQEMKAEERHOREKD 360
DB 301 ENMELAEERLKKLIEQYELREEHIDKVFPHKDLQOOLVDKLOQQAQEMKAEERHOREKD 360
QY 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWTK 420
DB 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWTK 420
QY 421 KIKKLEKETTMYRSRWSNKKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERND 480
DB 421 KIKKLEKETTMYRSRWSNKKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERND 480
QY 481 LNKRVQDLSAGGGSILDSGPERPBGCAQAPSSPRVTEAPCYPGAPSTASGQTGPQE 540
DB 481 LNKRVQDLSAGGGSILDSGPERPBGCAQAPSSPRVTEAPCYPGAPSTASGQTGPQE 540
QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 2
Q86T54
ID Q86T54 PRELIMINARY; PRT; 546 AA.
AC Q86T54;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKF2p451K215.
GN DKF2p451K215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832338; CAD91138.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728D674F78897DF6 CRC64;

Query Match 96.9%; Score 529; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKKNGAQSFKPSGPGPEAGPEGAQERPSQAAPAVEAEGSGSQAPKPEGAQA 60
DB 1 MKNQDKKNGAQSFKPSGPGPEAGPEGAQERPSQAAPAVEAEGSGSQAPKPEGAQA 60
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQEPAPPEPAEKSRITYVARN 120
DB 61 RTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQEPAPPEPAEKSRITYVARN 120
QY 121 GEPEPTVNGEKEPSKGPNTHEIRQSDVGVDRHRRPQEKKAAGLQKEITLLMQTLN 180
DB 121 GEPEPTVNGEKEPSKGPNTHEIRQSDVGVDRHRRPQEKKAAGLQKEITLLMQTLN 180
QY 181 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240
DB 181 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240

```

```

QY 241 LESLRELQRHNSLKEGVQARBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ 300
DB 241 LESLRELQRHNSLKEGVQARBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ 300
QY 301 ENMELAEERLKKLIEQYELREEHIDKVFPHKDLQOOLVDKLOQQAQEMKAEERHOREKD 360
DB 301 ENMELAEERLKKLIEQYELREEHIDKVFPHKDLQOOLVDKLOQQAQEMKAEERHOREKD 360
QY 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWTK 420
DB 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWTK 420
QY 421 KIKKLEKETTMYRSRWSNKKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERND 480
DB 421 KIKKLEKETTMYRSRWSNKKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERND 480
QY 481 LNKRVQDLSAGGGSILDSGPERPBGCAQAPSSPRVTEAPCYPGAPS 529
DB 481 LNKRVQDLSAGGGSILDSGPERPBGCAQAPSSPRVTEAPCYPGAPS 529

RESULT 3
Q86T86
ID Q86T86 PRELIMINARY; PRT; 546 AA.
AC Q86T86;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKF2p451I0918.
GN DKF2p451I0918.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832636; CAD89951.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; FBBCAD4F753F7F1 CRC64;

Query Match 79.3%; Score 433; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 13 QSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGSGSQAPKPEGAQARTAQSGALRDVS 72
DB 13 QSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGSGSQAPKPEGAQARTAQSGALRDVS 72
QY 73 BELSRQLEDILSTYCVNNQGGPGEDGAQEPAPPEPAEKSRITYVARNGEPEPTPVNGE 132
DB 73 BELSRQLEDILSTYCVNNQGGPGEDGAQEPAPPEPAEKSRITYVARNGEPEPTPVNGE 132
QY 133 KEPSKGPNTHEIRQSDVGVDRHRRPQEKKAAGLQKEITLLMQTLNLTSTPEKLAAL 192
DB 133 KEPSKGPNTHEIRQSDVGVDRHRRPQEKKAAGLQKEITLLMQTLNLTSTPEKLAAL 192
QY 193 CKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLSLCRELQRHN 252
DB 193 CKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLSLCRELQRHN 252
QY 253 RSLKEGVQARBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQENMELAEERLKKL 312
DB 253 RSLKEGVQARBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQENMELAEERLKKL 312
QY 313 IEQYELREEHIDKVFPHKDLQOOLVDKLOQQAQEMKAEERHOREKDFLLKEAVESORM 372
DB 313 IEQYELREEHIDKVFPHKDLQOOLVDKLOQQAQEMKAEERHOREKDFLLKEAVESORM 372
QY 373 CELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWTKCKKLEKETTMY 432

```

Db	373	CELKMQQETHLKOQALALYTEKFEFQNTLSKSEVFTTFBQEMKMTKKIKLEKETMY	432
Qy	433	RSRWESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVDLSAGG	492
Db	433	RSRWESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVDLSAGG	492
Qy	493	QGSLLTDSGPRRRPGCAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA	546
Db	493	QGSLLTDSGPRRRPGCAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA	546
RESULT 4			
ID	Q86T85	PRELIMINARY; PRT; 546 AA.	
AC	Q86T85		
DT	01-JUN-2003	(TREMELrel. 24, Created)	
DT	01-JUN-2003	(TREMELrel. 24, Last sequence update)	
DE	Hypothetical protein DKFZ451J0118.		
GN	DKFZ451J0118		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Pobo G., Han M.,		
RA	Osanger A., Wiemann S.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL832637; CA089922.1; -.		
KW	Hypothetical protein.		
SQ	SEQUENCE 546 AA; 61846 MW; 36718BAE3AA7E6C2 CRC64;		
Query Match 44.5%; Score 243; DB 4; Length 546;			
Best Local Similarity 99.5%; Pred. No. 8.8e-241;			
Matches 543; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	MKNQDKNGAAKQNPXSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEGAQA	60
Db	1	MKNQDKNGAAKQNPXSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEGAQA	60
Qy	61	RTAOSGALRDVSELSQLELDILSTYCVDNNGGPGDGAQGEPAEPAEDAEKSTYVARN	120
Db	61	RTAOSGALRDVSELSQLELDILSTYCVDNNGGPGDGAQGEPAEPAEDAEKSTYVARN	120
Qy	121	GEPEPTPVNGEKEPSPKDPNTEIRSDVGDHRRPQEKKAAGLGEITLLMOTLN	180
Db	121	GEPEPTPVNGEKEPSPKDPNTEIRSDVGDHRRPQEKKAAGLGEITLLMOTLN	180
Qy	181	TLSTPEEKLAALCKVAFLLAEHNSOKMKLLQKOSQOLVQEKDHLRGHSHKAVLARSK	240
Db	181	TLSTPEEKLAALCKVAFLLAEHNSOKMKLLQKOSQOLVQEKDHLRGHSHKAVLARSK	240
Qy	241	LESICRELQRHNRSLKEEGVQARAEERKEKXVTSHFQVTLNDIQLQMEQHNERNSKLQ	300
Db	241	LESICRELQRHNRSLKEEGVQARAEERKEKXVTSHFQVTLNDIQLQMEQHNERNSKLQ	300
Qy	301	ENMELAEELKKLIFQYELREEHIDVKFKHDLQOOLVDKLOQAQEMKEAREHOREKD	360
Db	301	ENMELAEELKKLIFQYELREEHIDVKFKHDLQOOLVDKLOQAQEMKEAREHOREKD	360
Qy	361	FLLEAVESQRCBIMKQOETHLQQAALALYTEKEFEFQNTLSKSEVFTTFKQEMERTK	420
Db	361	FLLEAVESQRCBIMKQOETHLQQAALALYTEKEFEFQNTLSKSEVFTTFKQEMERTK	420
Qy	421	KIKLEKETMYRSRWESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERND	480
Db	421	KIKLEKETMYRSRWESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERND	480
Qy	481	LNKRVDLSAGGQGSLLTDSGPRRRPGCAQAPSSPRVTEAPCYPGAPSTEASGOTGPQ	540
Db	481	LNKRVDLSAGGQGSLLTDSGPRRRPGCAQAPSSPRVTEAPCYPGAPSTEASGOTGPQ	540

Qy	541	PTSARA	546
Db	541	PTSARA	546
RESULT 5			
ID	Q86Y86	PRELIMINARY; PRT; 116 AA.	
AC	Q86Y86		
DT	01-JUN-2003	(TREMELrel. 24, Created)	
DT	01-JUN-2003	(TREMELrel. 24, Last sequence update)	
DE	Hypothetical protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	Strausberg R.;		
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC046565; AAH46565.1; -.		
KW	Hypothetical protein.		
SQ	SEQUENCE 116 AA; 12623 MW; B99B79EACAAB43F CRC64;		
Query Match 21.2%; Score 116; DB 4; Length 116;			
Best Local Similarity 100.0%; Pred. No. 1.1e-110;			
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	431	MYRSRWESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVDLSA	490
Db	1	MYRSRWESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVDLSA	60
Qy	491	GGGSLTDSGPRRRPGCAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA	546
Db	61	GGGSLTDSGPRRRPGCAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA	116
RESULT 6			
ID	Q8C5K1	PRELIMINARY; PRT; 150 AA.	
AC	Q8C5K1		
DT	01-MAR-2003	(TREMELrel. 23, Created)	
DT	01-MAR-2003	(TREMELrel. 23, Last sequence update)	
DE	Hypothetical protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	STRAIN=C57BL/6J; TISSUE=olfactory brain;		
RX	MEDLINE=22354683; PubMed=12466851;		
RA	The PANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs."		
RL	Nature 420:563-573 (2002).		
DR	EMBL; AK078192; BAC37168.1; -.		
KW	Hypothetical protein.		
FT	NON TER 150 150		
SQ	SEQUENCE 150 AA; 16490 MW; C444717503B42E49 CRC64;		
Query Match 3.8%; Score 21; DB 11; Length 150;			
Best Local Similarity 100.0%; Pred. No. 8.5e-13;			
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	75	LSRQLEDILSTYCVDNNGQSP	95
Db	66	LSRQLEDILSTYCVDNNGQSP	86

Db 97 LNTLSTPEKLAALCKKYA 115

RESULT 9

Q8BUK2 PRELIMINARY; PRT; 463 AA.

ID Q8BUK2

AC Q8BUK2; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE Muscle-derived protein MDP77 variant 1 (Fragment).

GN 2310001N14RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Heart;

RX MEDLINE=22354683; PubMed=12466851;

RA The PANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

RL EMBL; AK084639; BAC39238.1; -.

DR MGD; MGI:1916756; 2310001N14RIK.

FT NON TER 241 463

SQ SEQUENCE 463 AA; 54090 MW; 18E9F881D73D58AA CRC64;

Query Match 3.5%; Score 19; DB 11; Length 463;

Best Local Similarity 100.0%; Pred. No. 2.7e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AVLARKSLKSLCRLQHN 252

Db 210 AVLARKSLKSLCRLQHN 228

RESULT 10

Q8BHN1 PRELIMINARY; PRT; 524 AA.

ID Q8BHN1

AC Q8BHN1; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN RBBP7.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;

RX MEDLINE=22354683; PubMed=12466851;

RA The PANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

RL EMBL; AK030100; BAC26785.1; -.

DR EMBL; AK031783; BAC27547.1; -.

DR EMBL; AK044130; BAC31791.1; -.

DR MGD; MGI:1194910; Rbbp7.

DR GO; GO:0005834; C:nucleus; IDA.

DR GO; GO:0016564; P:negative regulation of transcription from P. . .; IDA.

DR GO; GO:0000122; P:negative regulation of transcription from P. . .; IDA.

KW Hypothetical protein.

SQ SEQUENCE 524 AA; 60308 MW; 0228777633B4ED7C CRC64;

Query Match 3.5%; Score 19; DB 11; Length 524;

Best Local Similarity 100.0%; Pred. No. 3e-10; Mismatches 0; Indels 0; Gaps 0;

Db 241 AA

RESULT 7

Q8BWJ7 PRELIMINARY; PRT; 241 AA.

ID Q8BWJ7

AC Q8BWJ7; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE Muscle-derived protein MDP77 variant 1 (Fragment).

GN 2310001N14RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Heart;

RX MEDLINE=22354683; PubMed=12466851;

RA The PANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

RL EMBL; AK052324; BAC34937.1; -.

DR MGD; MGI:1916756; 2310001N14RIK.

FT NON TER 241 241

SQ SEQUENCE 241 AA; 27232 MW; 0B2DD1BD215630BB CRC64;

Query Match 3.5%; Score 19; DB 11; Length 241;

Best Local Similarity 100.0%; Pred. No. 1.5e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AVLARKSLKSLCRLQHN 252

Db 210 AVLARKSLKSLCRLQHN 228

RESULT 8

Q8BP11 PRELIMINARY; PRT; 241 AA.

ID Q8BP11

AC Q8BP11; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN RBBP7.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;

RX MEDLINE=22354683; PubMed=12466851;

RA The PANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

RL EMBL; AK078477; BAC37296.1; -.

DR MGD; MGI:1194910; Rbbp7.

DR GO; GO:0005834; C:nucleus; IDA.

DR GO; GO:0016564; P:negative regulation of transcription from P. . .; IDA.

DR GO; GO:0000122; P:negative regulation of transcription from P. . .; IDA.

KW Hypothetical protein.

SQ SEQUENCE 241 AA; 27598 MW; 1E0C7A38E169F31A CRC64;

Query Match 3.5%; Score 19; DB 11; Length 241;

Best Local Similarity 100.0%; Pred. No. 1.5e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEKLAALCKKYA 197

	Matches	19;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	179	LNTLSTPEEKLAALCKKYA	197							
Db	146	LNTLSTPEEKLAALCKKYA	164							

RESULT 11

```

Q9NUQ3 ID Q9NUQ3 PRELIMINARY; PRT; 528 AA.
AC Q9NUQ3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ11209.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSU=Placenta;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Nanamiya K., Iwayanagi T.;
RT "VEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EXEL; AK002071; BAA92068.1; -
DR Genew; HGNC:18578; CXorf15.
KW Hypothetical protein.
SQ SEQUENCE 528 AA; 60605 MW; CAE8D781DE06ACE3 CRC64;

```

Query Match 3.5%; Score 19; DB 4; Length 528;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 19; Conservative 0; Mismatches 0; Indels

Qy 179 LNTLSTPEEKLAALCKKYA 197
Dd 146 LNTLSTPEEKLAALCKKYA 164

RESULT 12

Q8VBT1	PRELIMINARY;	PRT;	684 AA.
ID	Q8VBT1		
AC	Q8VBT1;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Muscle-derived protein MDP77 variant 2 (Muscle-derived protein MDP77 variant 1)		
DE	variant 1)		
GN	231000I1N14RIK OR MDP77.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
FX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Heart;		
RX	MEDLINE=21663722; PubMed=11805063;		
RA	Benson K.F., Chada K.;		
RA	"Molecular Characterization of the Mouse In(10)17Rk Inversion and		
RT	Identification of a Novel Muscle-Specific Gene at the Proximal		
RT	Breakpoint ";		
RL	Genetics 150:279-287(2002).		
DR	EMBL; AF422245; AAL33910.1; --		
DR	EMBL; AF422244; AAL33909.1; --		
DR	MG; MG1:1916756; 231000I1N14RIK.		
SQ	SEQUENCE 684 AA; 77049 MW; EBA64E28EEF56E31 CRC64;		

```

Query Match      3.5%; Score 19; DB 11; Length 684;
Best Local Similarity 100.0%; Pred No. 3.8e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 234 AVLARSKLESKLELQRNH 252
      |||||
Db 209 AVLARSKLESKLELQRNH 227
      |||||

```

RESULT 13

Q9P0X1	PRELIMINARY;	PRT;	186 AA.
AD	Q9P0X1		
IC	Q9P0X1;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-WAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Lipopolysaccharide specific response-5 protein (Fragment).		
DE	LSR5		
GN	Homo sapiens (Human)		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
OC	[1]		
RN	SEQUENCE FROM N.A.		
RP	Chai Y.B., Zhao Z.L., Zhu P., Yan W., Chen N.C., Wang Q., Yue L.,		
RA	Chen S.M.;		
RA	"New Homo sapiens gene from dental pulp cells."		
RT	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AF143740; AAF70546.2; -;		
RL	NON TER		
SC	1		
FT	SEQUENCE	186 AA; 21464 MW; 93BF83838096B0A CRC64;	

Query Match 3.1%; Score 17; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels

Qy 413 QEMEKNTKKIKKLEKET 429
|||
Db 38 QEMEKNTKKIKKLEKET 54

RESULT 14

Q8N3S2	PRELIMINARY;	PRT;	505 AA.
AC	Q8N3S2		
AD	Q8N3S2;		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)		
DE	Hypothetical protein (Fragment).		
DE	Hypothetical protein (Fragment).		
GN	DKFZP451A175.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
XP	[1]		
RP	SEQUENCE FROM N.A.		
RA	Koehler K., Beyer A., Mewes H.W., Weil B., Wiemann S.;		
RA	Submitted (JUL1/2002) to the EMBL/GenBank/DBDJB databases.		
RA	EMBL; AL832322; CAD38617.1; -.		
DR	Hypothetical protein.		
DR	Hypothetical protein.		
FT	NC TER 1		
FT	NC TER 1		
SO	SEQUENCE 505 AA; 56957 MW; 1AA33548A2F1DADE CRC64;		

Query Match	3.1%;	Score 17;	DB 4;	Length 505;
Best Local Similarity	100.0%;	Pred. No. 3.3e-08;		
Matches	17;	Conservative	0;	Mismatches
			0;	Indels

QY 236 LARSKLESLORELOH 252
 236 LARSKLESLORELOH 252
 30 LARSKLESLORELOH 46
 30 LARSKLESLORELOH 46

RESULT 15

CT TRACKING

Q9I969
ID Q9I969 PRELIMINARY; PRT; 676 AA.
AC Q9I969;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscle derived protein.
GN MDP77.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=20175243; PubMed=10708594;
RA Uyeda A.; Fukui I.; Fujimori K.; Kiyosue K.; Nishimune H.; Kasai M.;
RA Taguchi T.;
RT "MDP77: A novel neurite-outgrowth-promoting protein predominantly
RT expressed in chick muscles.";
RL Biochem. Biophys. Res. Commun. 269:564-569(2000).
DR EMBL; D89999; BAA94755.1; -.
DR FIR; JC7222; JC7222.
SQ SEQUENCE 676 AA; 77020 MW; FCEA9E393250EE94 CRC64;

Query Match 3.1%; Score 17; DB 13; Length 676;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 LARSKLESICRELOPHN 252
DB 207 LARSKLESICRELOPHN 223
|||||
|||||

Search completed: June 8, 2004, 16:39:31
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:36:52 ; Search time 22 Seconds
(without alignments)
1281.263 Million cell updates/sec

Title: US-10-023-529-44
Perfect score: 546
Sequence: 1 MNQDKNGAKQSNPKSP.....APSTRASGQTGPQETISARA 546

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/53 COMB.pep.*
3: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/53 COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	546	100.0	546	4	US-09-616-289-44
2	429	78.6	530	4	US-08-979-608A-8
3	429	78.6	530	4	US-09-517-849-8
4	429	78.6	530	4	US-09-616-289-8
5	172	31.5	557	4	US-08-979-608A-5
6	172	31.5	557	4	US-09-517-849-5
7	172	31.5	557	4	US-09-616-289-5
8	19	3.5	386	2	US-08-968-751-6
9	8	1.5	185	4	US-09-352-991A-19490
10	8	1.5	260	2	US-08-879-561-7
11	7	1.3	7	4	US-08-979-608A-41
12	7	1.3	7	4	US-09-517-849-41
13	7	1.3	7	4	US-09-616-289-41
14	7	1.3	64	4	US-09-489-039A-9565
15	7	1.3	77	3	US-09-146-950-25
16	7	1.3	105	4	US-08-732-210-434
17	7	1.3	124	4	US-09-489-039A-10493
18	7	1.3	127	4	US-08-849-303-19
19	7	1.3	142	4	US-09-489-039A-9418
20	7	1.3	159	3	US-09-146-950-20
21	7	1.3	172	4	US-09-079-030-117
22	7	1.3	187	4	US-09-328-352-7281
23	7	1.3	197	3	US-09-146-950-18
24	7	1.3	211	4	US-09-434-001C-5175
25	7	1.3	261	2	US-08-879-561-1
26	7	1.3	267	4	US-09-489-039A-13457
27	7	1.3	274	1	US-08-287-959-3

28	7	1.3	277	4	US-09-857-498A-7	Sequence 7, Appli
29	7	1.3	281	4	US-09-134-000C-3980	Sequence 3980, Ap
30	7	1.3	294	4	US-09-489-039A-8327	Sequence 8327, Ap
31	7	1.3	302	4	US-09-328-352-4508	Sequence 4508, Ap
32	7	1.3	311	4	US-09-252-991A-28792	Sequence 28792, A
33	7	1.3	312	1	US-08-285-440-4	Sequence 4, Appli
34	7	1.3	312	1	US-08-630-349-4	Sequence 4, Appli
35	7	1.3	314	4	US-09-252-991A-18599	Sequence 18599, A
36	7	1.3	330	4	US-09-252-991A-28531	Sequence 28531, A
37	7	1.3	341	3	US-09-008-465-1	Sequence 1, Appli
38	7	1.3	341	4	US-09-528-959-1	Sequence 1, Appli
39	7	1.3	346	4	US-09-252-991A-27167	Sequence 27167, A
40	7	1.3	364	4	US-09-252-991A-27209	Sequence 27209, A
41	7	1.3	388	4	US-09-252-991A-21572	Sequence 21572, A
42	7	1.3	411	4	US-09-252-991A-24632	Sequence 24632, A
43	7	1.3	422	4	US-09-252-991A-21571	Sequence 21571, A
44	7	1.3	446	4	US-09-252-991A-25318	Sequence 25318, A
45	7	1.3	473	4	US-09-252-991A-16904	Sequence 16904, A

ALIGNMENTS

RESULT 1
US-09-616-289-44
; Sequence 44, Application US/09616289
; Patent No. 5632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-616-289-44

Query Match		100.0%;	Score 546;	DB 4;	Length 546;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 546;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKNQDKNGAKQSNPKSPQOPEAGPEGAGQERPSQAAPAVEAEFGSGSQAPRKPEGQAQ	60		
DB	1	MKNQDKNGAKQSNPKSPQOPEAGPEGAGQERPSQAAPAVEAEFGSGSQAPRKPEGQAQ	60		
QY	61	RTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEAGQCEPAEPDAKSRITVARN	120		
DB	61	RTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEAGQCEPAEPDAKSRITVARN	120		
QY	121	GRPEPTPVNGEKEPSKGDPMTEETIRQSDVGDHRDRRPOEKKKAGLGKEITLLMQTLN	180		
DB	121	GRPEPTPVNGEKEPSKGDPMTEETIRQSDVGDHRDRRPOEKKKAGLGKEITLLMQTLN	180		
QY	181	TLSTPEEKLAALCKKYAEILLEHRNSQOMKLLQKQSQLVQEKDHLRGEHSKAVLARSK	240		
DB	181	TLSTPEEKLAALCKKYAEILLEHRNSQOMKLLQKQSQLVQEKDHLRGEHSKAVLARSK	240		

```

QY 241 LESLRELRNRSLKBEVGQVRAAREBEKREKXVTSFQVTLNDIOLQMEQHNRNSKLRLQ 300
DB 241 LESLRELRNRSLKBEVGQVRAAREBEKREKXVTSFQVTLNDIOLQMEQHNRNSKLRLQ 300
QY 301 ENMELARLKLIEQYELREHIDKVPKHKDLOQOLVDKLOQAOQEMLKEAEERHOREKD 360
DB 301 ENMELARLKLIEQYELREHIDKVPKHKDLOQOLVDKLOQAOQEMLKEAEERHOREKD 360
QY 361 FLLEAVESQRMCELMKQETHLQKQALALYTEKEFEFQNTLSKSEVFTTFKQEMERQTK 420
DB 361 FLLEAVESQRMCELMKQETHLQKQALALYTEKEFEFQNTLSKSEVFTTFKQEMERQTK 420
QY 421 KIKLEKETTMYRSWESSNALLEMAEKTVRKLEGLQVKIORLEKLCRALQTERND 480
DB 421 KIKLEKETTMYRSWESSNALLEMAEKTVRKLEGLQVKIORLEKLCRALQTERND 480
QY 481 LNKRVQDLASAGQSGSLTDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQ 540
DB 481 LNKRVQDLASAGQSGSLTDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQ 540
QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 2
US-08-979-608A-8
; Sequence 8, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;               Lees, Robert S.
;               Law, Simon W.
;               Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                   TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```

```

US-08-979-608A-8
Query Match 78.6%; Score 429; DB 4; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 KSPGCGPEAGPEGAQPRPSQAPPAVEAEGPGSGQAPRKPEGAQARTAQSGALRDVSEELS 76
DB 1 KSPGCGPEAGPEGAQPRPSQAPPAVEAEGPGSGQAPRKPEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEEDILSTVCVDNNQGGGEGDGAQGEPAEPDAEKSRTTYVARNGEPPEPTPVVNGEKEPS 136
DB 61 RQLEEDILSTVCVDNNQGGGEGDGAQGEPAEPDAEKSRTTYVARNGEPPEPTPVVNGEKEPS 120
QY 137 KGPNTTEERQSDVEGDRDRRPOBKKAAGLQKBEITLLMOTLNTLTSTPEEKLAALCKKY 196
DB 121 KGPNTTEERQSDVEGDRDRRPOBKKAAGLQKBEITLLMOTLNTLTSTPEEKLAALCKKY 180
QY 197 AELEBEHRNSQKMLLQKKQSQOLVQEKDHLRGEHSKAVLARSKLRLSLCEZLQHRNRSK 256
DB 181 AELEBEHRNSQKMLLQKKQSQOLVQEKDHLRGEHSKAVLARSKLRLSLCEZLQHRNRSK 240
QY 257 EBGVQARBEERKEKVTSHFQVTLNDIOLQMEQHNRNSKLROENNELARLKLIEQY 316
DB 241 EBGVQARBEERKEKVTSHFQVTLNDIOLQMEQHNRNSKLROENNELARLKLIEQY 300
QY 317 ELREEHIDKVPKHKDLOQOLVDKLOQAOQEMLKEAEERHOREKDFLLEAVESQRMCELM 376
DB 301 ELREEHIDKVPKHKDLOQOLVDKLOQAOQEMLKEAEERHOREKDFLLEAVESQRMCELM 360
QY 377 KQETHLKOALALYTEKEFEFQNTLSKSEVFTTFKQEMERQTKIKLEKETTTMYRSRW 436
DB 361 KQETHLKOALALYTEKEFEFQNTLSKSEVFTTFKQEMERQTKIKLEKETTTMYRSRW 420
QY 437 ESSNKALLEMAEKTVRKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLASAGQSGSL 496
DB 421 ESSNKALLEMAEKTVRKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLASAGQSGSL 480
QY 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQPTSARA 546
DB 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQPTSARA 530

RESULT 3
US-09-517-849-8
; Sequence 8, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;               Lees, Robert S.
;               Law, Simon W.
;               Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                   TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608

```

FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-517-849-8

Query Match 78.6%; Score 429; DB 4; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSQAPRKEGAQARTAQSGALRDVSEELS 76
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSQAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KQDPNTEIRQSDVEGDRHRRPOEKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY 196
DB 121 KQDPNTEIRQSDVEGDRHRRPOEKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
QY 197 AELLBEHRNSQKMKLQKQSQQLVQEKDLRGEHSKAVLARSKLESCLRELQHRNRSK 256
DB 181 AELLBEHRNSQKMKLQKQSQQLVQEKDLRGEHSKAVLARSKLESCLRELQHRNRSK 240
QY 257 EGVQARAREEKEKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 316
DB 241 EGVQARAREEKEKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
QY 317 ELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWKIKKLEKETTWYRSRW 436
DB 361 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWKIKKLEKETTWYRSRW 420
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSL 496
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSL 480
QY 497 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 546
DB 481 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530

FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-517-849-8

Query Match 78.6%; Score 429; DB 4; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSQAPRKEGAQARTAQSGALRDVSEELS 76
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSQAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KQDPNTEIRQSDVEGDRHRRPOEKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY 196
DB 121 KQDPNTEIRQSDVEGDRHRRPOEKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
QY 197 AELLBEHRNSQKMKLQKQSQQLVQEKDLRGEHSKAVLARSKLESCLRELQHRNRSK 256
DB 181 AELLBEHRNSQKMKLQKQSQQLVQEKDLRGEHSKAVLARSKLESCLRELQHRNRSK 240
QY 257 EGVQARAREEKEKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 316
DB 241 EGVQARAREEKEKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
QY 317 ELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWKIKKLEKETTWYRSRW 436
DB 361 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWKIKKLEKETTWYRSRW 420
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSL 496
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSL 480
QY 497 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 546
DB 481 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530

RESULT 5
US-08-979-608A-5
Sequence 5, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 225 Franklin Street

RESULT 4
US-09-517-849-8
Sequence 8, Application US/09616289
Patent No. 6632523
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/616,289
CURRENT FILING DATE: 2000-07-14

```
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

Query Match 31.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 5.2e-150;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 TLNDIQLQMEQHNRNSKLRQENMELARLKLIEQYELREEHIDKVFKHKLQQLQVLDA 339
DB 280 TLNDIQLQMEQHNRNSKLRQENMELARLKLIEQYELREEHIDKVFKHKLQQLQVLDA 339
QY 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
DB 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
QY 400 TLSKSSEVFTTFKQEMERTKIKKLEKETTMYRSRWSSNKALLEABEKT 451
DB 400 TLSKSSEVFTTFKQEMERTKIKKLEKETTMYRSRWSSNKALLEABEKT 451

RESULT 6
US-09-517-849-5
; Sequence 5, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-517-849-5

Query Match 31.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 5.2e-150;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 TLNDIQLQMEQHNRNSKLRQENMELARLKLIEQYELREEHIDKVFKHKLQQLQVLDA 339
DB 280 TLNDIQLQMEQHNRNSKLRQENMELARLKLIEQYELREEHIDKVFKHKLQQLQVLDA 339
QY 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
DB 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
QY 400 TLSKSSEVFTTFKQEMERTKIKKLEKETTMYRSRWSSNKALLEABEKT 451
DB 400 TLSKSSEVFTTFKQEMERTKIKKLEKETTMYRSRWSSNKALLEABEKT 451

RESULT 7
US-09-616-289-5
; Sequence 5, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-616-289-5
```


Query Match 31.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 5.2e-150;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 TLNDIQLOMPOHNRNSKLRQENMELARLKLLIEQYELAREHIDKVPKHDIQQQLVDA 339
|||||
Db 280 TLNDIQLOMPOHNRNSKLRQENMELARLKLLIEQYELAREHIDKVPKHDIQQQLVDA 339
|||||

Qy 340 KLAQAQMLKEABERHQRKDFLLKEAVESQRMCELMKQOETHLKQOLALYTKPFQFN 399
|||||
Db 340 KLAQAQMLKEABERHQRKDFLLKEAVESQRMCELMKQOETHLKQOLALYTKPFQFN 399
|||||

Qy 400 TSKSSSVFTFKQEKEMTKKKLKEKETTMYRSRWESSNKALLEAREKT 451
|||||
Db 400 TSKSSSVFTFKQEKEMTKKKLKEKETTMYRSRWESSNKALLEAREKT 451
|||||

RESULT 8
US-08-968-751-6
; Sequence 6, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfeld, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Tetyln T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,751
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Giotta, Gregory
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX1024 GG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 262-8710
; TELEFAX: (510) 222-9758
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-968-751-6

Query Match 3.5%; Score 19; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 INTLSTPEEKLAALCKKYA 197
|||||
Db 135 INTLSTPEEKLAALCKKYA 153
|||||

RESULT 9
US-09-252-991A-19490
; Sequence 19490, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19490
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19490

Query Match 1.5%; Score 8; DB 4; Length 165;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 DRDHRPQ 160
|||||
Db 40 DRDHRPQ 47
|||||

RESULT 10
US-08-879-561-7
; Sequence 7, Application US/08879561
; Patent No. 5817482
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/879,561
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0325 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

```
; LIBRARY: GenBank
; CLONE: 1480198
US-08-879-561-7

Query Match      1.5%; Score 8; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 261 QARAREEE 268
Db 188 QARAREEE 195
|||||

RESULT 11
US-08-979-608A-41
; Sequence 41, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-08-979-608A-41

Query Match      1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 RDVSEEL 75
Db 1 RDVSEEL 7
|||||

RESULT 13
US-09-616-289-41
; Sequence 41, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
```

QY 161 EKKKXKG 167
 Db 71 EKKKXKG 77
 Search completed: June 8, 2004, 16:40:36
 Job time : 30 secs

PRIOR APPLICATION NUMBER: US 60/031,930
 PRIOR FILING DATE: 1996-11-27
 PRIOR APPLICATION NUMBER: US 60/048,547
 PRIOR FILING DATE: 1997-06-03
 NUMBER OF SEQ ID NOS: 53
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 41
 LENGTH: 7
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-616-289-41

Query Match 1.3%; Score 7; DB 4; Length 7;
 Best Local Similarity 100.0%; Pred. No. 38+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 RDVSEEL 75
 Db 1 RDVSEEL 7

RESULT 14
 US-09-489-039A-9565
 Sequence 9565, Application US/09489039A
 Patent No. 6610836
 GENERAL INFORMATION:
 APPLICANT: Gary Breton et. al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 2709.2004001
 CURRENT APPLICATION NUMBER: US/09/489,039A
 CURRENT FILING DATE: 2000-01-27
 PRIOR APPLICATION NUMBER: US 60/117,747
 PRIOR FILING DATE: 1999-01-29
 NUMBER OF SEQ ID NOS: 14342
 SEQ ID NO 9565
 LENGTH: 64
 TYPE: PRT
 ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-9565

Query Match 1.3%; Score 7; DB 4; Length 64;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VVNGEKE 134
 Db 39 VVNGEKE 45

RESULT 15
 US-09-146-950-25
 Sequence 25, Application US/09146950A
 Patent No. 6287808
 GENERAL INFORMATION:
 APPLICANT: Busfield, Samantha J.
 TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
 TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
 FILE REFERENCE: 09404/057001
 CURRENT APPLICATION NUMBER: US/09/146,950A
 CURRENT FILING DATE: 1998-09-03
 NUMBER OF SEQ ID NOS: 25
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 25
 LENGTH: 77
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-146-950-25

Query Match 1.3%; Score 7; DB 3; Length 77;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:33:31 ; Search time 60 Seconds
(without alignments)
2571.182 Million cell updates/sec

Title: US-10-023-529-44
Perfect score: 546
Sequence: 1 MKNQDKKNGAAKQSNPKSP.....APSTASGGTGQPEPTSARA 546

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1586107 seqs, 282547505 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

- Database : A Geneseq_29Jan04:*
- 1: Geneseqp1980s:*
 - 2: Geneseqp1990s:*
 - 3: Geneseqp2000s:*
 - 4: Geneseqp2001s:*
 - 5: Geneseqp2002s:*
 - 6: Geneseqp2003as:*
 - 7: Geneseqp2003bs:*
 - 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	546	100.0	546	4 AAB82808	Aab82808 Human low
2	546	100.0	639	6 ABR41210	AbR41210 Human DIT
3	546	100.0	639	6 ABR41186	AbR41186 Human DIT
4	476	87.2	510	4 ABB11764	Abb11764 Human LDL
5	476	87.2	510	4 AAM79741	Aam79741 Human pro
6	429	78.6	530	2 AAW49042	Aaw49042 Human low
7	429	78.6	530	4 AAB82804	Aab82804 Human low
8	356	65.2	356	6 ABO07223	AbO07223 Human p53
9	347	63.6	358	4 AAM78757	Aam78757 Human pro
10	228	41.8	1749	4 ABG00839	Abg00839 Novel hum
11	172	31.5	557	2 AAW49039	Aaw49039 Rabbit lo
12	172	31.5	557	4 AAB82801	Aab82801 Rabbit lo
13	139	3.5	204	4 AAB63259	Aab63259 Human bre
14	139	3.5	386	2 AAW37883	Aaw37883 BRCA1 mod
15	139	3.5	386	2 AAY30151	Aay30151 Amino aci
16	139	3.5	395	6 ABO07222	AbO07222 Human p53
17	139	3.5	528	4 AAB93594	Ab93594 Human pro
18	139	3.1	85	5 ABP42281	Abp42281 Human ova
19	139	3.1	676	2 AAY17863	Aay17863 Neurite e
20	139	3.1	684	6 ABR43284	AbR43284 Human neu
21	12	2.2	475	4 AAO08301	Aao08301 Human pol
22	12	2.2	475	7 ADC33216	Adc33216 Human nov
23	9	1.6	99	4 AAU23308	Aau23308 Human car
24	9	1.6	99	7 ADE46276	Ade46276 Human car
25	9	1.6	554	4 ABB59454	Abb59454 Drosophill

26	9	1.6	637	5 ABU44191	Abu44191 Protein e
27	9	1.6	729	2 AAR89275	Aar89275 Yeast coa
28	8	1.5	65	2 AAY12665	Aay12665 Human 5'
29	8	1.5	156	2 AAY01195	Aay01195 Polypteti
30	8	1.5	170	3 AAG08727	Aag08727 Arabidops
31	8	1.5	174	3 AAG08726	Aag08726 Arabidops
32	8	1.5	194	4 ABG19843	Abg19843 Novel hum
33	8	1.5	224	3 AAG20570	Aag20570 Arabidops
34	8	1.5	229	5 ABU05565	Abu05565 M. tuberc
35	8	1.5	360	5 ABG77243	Abg77243 Selected
36	8	1.5	360	5 ABU11114	Abj11114 Yeast sel
37	8	1.5	364	5 ABJ11057	Abj11057 Yeast sel
38	8	1.5	388	7 ADB95026	Adb95026 A. thalia
39	8	1.5	416	7 ADC31613	Adc31613 Human nov
40	8	1.5	484	4 AAB79356	Aab79356 Corynebac
41	8	1.5	484	4 AAG92676	Aag92676 C. glutami
42	8	1.5	785	3 AAG41279	Aag41279 Arabidops
43	8	1.5	823	3 AAG41278	Aag41278 Arabidops
44	8	1.5	836	3 AAG41277	Aag41277 Arabidops
45	8	1.5	952	7 ADB95084	Adb95084 A. thalia

ALIGNMENTS

RESULT 1
AAB82808
ID AAB82808 standard; protein; 546 AA.
XX
AC AAB82808;
XX
DT 12-NOV-2001 (first entry)
XX
DE Human low density lipoprotein binding protein 3 (LBP-3).
XX
XX Low density lipoprotein binding protein 3; LBP-3; LDL; human;
XX atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
OS Homo sapiens.
XX
PN WO200164874-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006355.
XX
PR 02-MAR-2000; 2000US-00517849.
PR 14-JUL-2000; 2000US-00616289.
XX (BOST-) BOSTON HEART FOUND INC.
PI Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 2001-565505/63.
XX N-PSDB; AAH26501.
XX New isolated low density lipoprotein binding polypeptide for treating,
diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX Claim 13(1); Fig 8A; 143pp; English.

The present sequence is that of novel human low density lipoprotein binding protein 3 (LBP-3). The amino acid sequence was deduced from the coding region of isolated genomic DNA (see AAH26501). It differs from the sequence predicted from an LBP-3 cDNA clone (see AAB82804) by the presence of an additional 16 amino acids at the N-terminus (the cDNA clone is 5' truncated) and by having asparagine at amino acid position 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP polypeptides of the invention that are capable of binding to native and methylated low density lipoproteins. Also claimed are biologically active fragments and analogues of these LBPs, polynucleotides encoding LBPs, as well as expression vectors, cells and methods of producing the LBPs. A polypeptide having the amino acid residues 96-110 of the present sequence

CC is claimed (see AAB82820). Methods of determining if an animal is at risk
 CC for atherosclerosis, methods for evaluating an agent for use in treating
 CC atherosclerosis, and methods for treating a cell having an abnormality in
 CC structure or metabolism of LBP are claimed. Pharmaceutical compositions
 CC comprising an LBP polypeptide or nucleic acid, and vaccine compositions,
 CC are also claimed

XX
 SQ Sequence 546 AA;

Query Match 100.0%; Score 546; DB 4; Length 546;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNQDKKNGAAKQSNPKSPGQPEAGPGAGPQQAAPAVAEAGPGSSQAPRPEGAQA 60

Db 1 MNQDKKNGAAKQSNPKSPGQPEAGPGAGPQQAAPAVAEAGPGSSQAPRPEGAQA 60

Qy 61 RTAQSGALRDVSEELSRQLEDILSTYCVDDNQGGPGEAGQCEPAEPDAEKSTYVARN 120

Db 61 RTAQSGALRDVSEELSRQLEDILSTYCVDDNQGGPGEAGQCEPAEPDAEKSTYVARN 120

Qy 121 GEPEPTPVVNGEKEPSKGDPTETIRQSDVGRDHRPQEKKAKGKGTITLLMOTLN 180

Db 121 GEPEPTPVVNGEKEPSKGDPTETIRQSDVGRDHRPQEKKAKGKGTITLLMOTLN 180

Qy 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSQQLVQEKDHLRGEHSKAVLARSK 240

Db 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSQQLVQEKDHLRGEHSKAVLARSK 240

Qy 241 LESLCRELQRNRSLEKGVQAPAREEKKKEVTSHPQVTLNDIQLQEQNEENSKLRQ 300

Db 241 LESLCRELQRNRSLEKGVQAPAREEKKKEVTSHPQVTLNDIQLQEQNEENSKLRQ 300

Qy 301 ENVELAERLKKLIEQYELREEDIKVFHKQIQQLVDKQLQQAQEMLKEAEERHQREKD 360

Db 301 ENVELAERLKKLIEQYELREEDIKVFHKQIQQLVDKQLQQAQEMLKEAEERHQREKD 360

Qy 361 FLKKEAVESQRCMLKQOETHLKOALYTKPEEPQNTLSKSEVTTTKQEMEKTK 420

Db 361 FLKKEAVESQRCMLKQOETHLKOALYTKPEEPQNTLSKSEVTTTKQEMEKTK 420

Qy 421 KIKKLEKETTWYRSWESSNKALLEMAEKTVDKELSGLOVKIQRLKLCRALQTERND 480

Db 421 KIKKLEKETTWYRSWESSNKALLEMAEKTVDKELSGLOVKIQRLKLCRALQTERND 480

Qy 481 LNKRVODLSAGQGSILTDSGPERRPREGCAQAPSPRVTEAPCTYCPGAPSTASGQTGPQE 540

Db 481 LNKRVODLSAGQGSILTDSGPERRPREGCAQAPSPRVTEAPCTYCPGAPSTASGQTGPQE 540

Qy 541 PTSARA 546

Db 541 PTSARA 546

RESULT 2

ID ABR41210

XX ABR41210 standard; protein; 639 AA.

AC ABR41210;

XX ABR41210;

XX 02-JUN-2003 (first entry)

DT Human DITHP extracellular signalling protein.

DE Human DITHP extracellular signalling protein.

XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;

KW cancer; cell proliferative disorder; autoimmune disorder;

KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;

KW neurological disorder; gastrointestinal disorder; transport disorder;

KW connective tissue disorder; drug screening; proteome analysis;

KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;

KW disease model; toxicological testing; transcript imaging;

KW extracellular signalling.

XX

OS Homo sapiens.

XX W0200297031-A2.

XX PN

XX PD 05-DEC-2002.

XX XX

XX 27-MAR-2002; 2002MO-US010056.

XX XX

XX 28-MAR-2001; 2001US-0279619P.

XX 29-MAR-2001; 2001US-0280067P.

XX 29-MAR-2001; 2001US-0280068P.

XX 16-MAY-2001; 2001US-0291280P.

XX 17-MAY-2001; 2001US-0291829P.

XX 17-MAY-2001; 2001US-0291849P.

XX 19-JUN-2001; 2001US-0299428P.

XX 20-JUN-2001; 2001US-0299776P.

XX 20-JUN-2001; 2001US-0300001P.

XX XX

XX (INCY-) INCYTE GENOMICS INC.

XX XX

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;

XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

XX Peralta Ch, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-129518/12.

XX N-PSDB; ACC46153.

XX Novel human diagnostic and therapeutic polypeptide useful for identifying

XX test compound which specifically binds to a polypeptide encoded by human

XX diagnostic and therapeutic polynucleotide, and to induce antibodies.

XX Claim 27; SEQ ID NO 745; 591pp; English.

XX The invention relates to novel human diagnostic and therapeutic

XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded

XX proteins (DITHP; ABR41136-ABR41812). The invention also relates to

XX polynucleotide sequences at least 90% identical to the dithp cDNA

XX sequences of the invention; recombinant vectors, host cells and

XX transgenic organisms comprising a dithp nucleic acid sequence; the

XX recombinant production of DITHP proteins; antibodies specific for DITHP

XX proteins; microarrays comprising dithp nucleic acid sequences; methods of

XX detecting dithp nucleotide and protein sequences; methods of screening

XX for compounds which specifically bind a DITHP protein; and methods of

XX assessing the toxicity of test compounds using a dithp hybridisation

XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the

XX diagnosis of a wide variety of conditions including cancer and other cell

XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,

XX viral, fungal or parasitic infections; hormonal disorders; metabolic

XX disorders; neurological disorders; gastrointestinal disorders; transport

XX disorders; and connective tissue disorders. They may also be used to

XX screen for modulators of protein activity or gene expression. DITHP

XX proteins can additionally be used in analysis of the proteome of a tissue

XX or cell type and to induce antibodies. The dithp nucleic acids are

XX additionally useful in somatic or germline gene therapy of the disorders

XX mentioned above, as a source of antisense sequences, as a source of

XX probes and primers, in genotyping and identification of individuals, in

XX the generation of transgenic animal models of human disease or knock in

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 100.0%; Score 546; DB 6; Length 639;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNQDKKNGAAKQSNPKSPGQPEAGPGAGPQQAAPAVAEAGPGSSQAPRPEGAQA 60

|||||

PR	20-JUN-2001; 2001US-0299776P.
PR	20-JUN-2001; 2001US-0300001P.
XX	(INCY-) INCYTE GENOMICS INC.
PI	Daffo A, Jones AL, Tran AB, Bahl CR, Gietzen D, Chinn J;
PI	Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Agnew SR;
PI	Daugherty SC, Dam TC, Liu TP, Nguyen Y, Gerstin EH;
PI	Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI	Flores V, Marwaha R, Lo A, Lan RY, Urashka MB;
XX	WPI; 2003-129518/12.
DR	N-PSDB; ACC46130.
XX	Novel human diagnostic and therapeutic polypeptide useful for identifying
PT	test compound which specifically binds to a polypeptide encoded by human
PT	diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX	Claim 27; SEQ ID NO 721; 591pp; English.
PS	The invention relates to novel human diagnostic and therapeutic
XX	polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC	proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC	polynucleotide sequences at least 90% identical to the dithp cDNA
CC	sequences of the invention; recombinant vectors, host cells and
CC	transgenic organisms comprising a dithp nucleic acid sequence; the
CC	recombinant production of DITHP proteins; antibodies specific for DITHP
CC	proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC	detecting dithp nucleotide and protein sequences; methods of screening
CC	for compounds which specifically bind a DITHP protein; and methods of
CC	assessing the toxicity of test compounds using a dithp hybridisation
CC	probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC	diagnosis of a wide variety of conditions including cancer and other cell
CC	proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC	viral, fungal or parasitic infections; hormonal disorders; metabolic
CC	disorders; neurological disorders; gastrointestinal disorders; transport
CC	disorders; and connective tissue disorders. They may also be used to
CC	screen for modulators of protein activity or gene expression. DITHP
CC	proteins can additionally be used in analysis of the proteome of a tissue
CC	or cell type and to induce antibodies. The dithp nucleic acids are
CC	additionally useful in somatic or germline gene therapy of the disorders
CC	mentioned above, as a source of antisense sequences, as a source of
CC	probes and primers, in genotyping and identification of individuals, in
CC	the generation of transgenic animal models of human disease or knock in
CC	humanised animals, in toxicological testing, and in transcript imaging.
CC	The present sequence represents a DITHP protein which has extracellular
CC	signalling activity. Note: The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 639 AA;
SQ	Query Match 100.0%; Score 546; DB 6; Length 639;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MKNQDKNGAAKQSNPKSPGQPEAGPEGAGPQAPPAVEAGPGSSQAPRKPEGAQA 60
DB	94 MKNQDKNGAAKQSNPKSPGQPEAGPEGAGPQAPPAVEAGPGSSQAPRKPEGAQA 153
QY	61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAGPAPDAEKSRITYARN 120
DB	154 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAGPAPDAEKSRITYARN 213
QY	121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVGRDHRHPOEKKAKGLGKEITLLMOTLN 180
DB	214 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVGRDHRHPOEKKAKGLGKEITLLMOTLN 273
QY	181 TLSTPEEKLAALCKKYAELLEERHNSQKMLQKQSQLVQKQHLRGEHSAVLARSK 240
DB	274 TLSTPEEKLAALCKKYAELLEERHNSQKMLQKQSQLVQKQHLRGEHSAVLARSK 333
QY	241 LESICRELQHRNLSKEEGVQARSEEEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQR 300

Db 432 KIKLEKETTMYRSWESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQT 487

RESULT 5

AAW79741

ID AAW79741 standard; protein; 510 AA.

AC AAW79741;

XX 06-NOV-2001 (first entry)

DT Human protein SEQ ID NO 3387.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

FN 09-AUG-2001.

PD 05-FEB-2001; 2001NO-US004098.

XX 03-FEB-2000; 2000US-00495914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI: 2001-476283/51.

DR N-PSDB; AAK52874.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.

PT Claim 20; Page 319; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

XX Sequence 510 AA;

Query Match 87.2%; Score 476; DB 4; Length 510;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKNGAAKQSNPKSSPGPEAGPEGAQEPQAPVAEAGCGSSQAPKPEGAQA 60

Db 12 MKNQDKNGAAKQSNPKSSPGPEAGPEGAQEPQAPVAEAGCGSSQAPKPEGAQA 71

QY 61 RTAQSGALRDVSELSROLEDILSTYCVNNDNQGPGEDGAQGEPAEPDAKSKRTTYARN 120

Db 72 RTAQSGALRDVSELSROLEDILSTYCVNNDNQGPGEDGAQGEPAEPDAKSKRTTYARN 131

QY 121 GEPEPTFWNGEKEPSKGDPTNTEBIROSDEVDGDRHRRPOEKKAKAGLGK3ITILLMOTLN 180

Db 132 GEPEPTFWNGEKEPSKGDPTNTEBIROSDEVDGDRHRRPOEKKAKAGLGK3ITILLMOTLN 191

QY 181 TLSTPEEKLAALCKKYAELLLEHNSQKMLLQKQSQOLVOEKDHLGEGSKAVIARSK 240

Db 192 TLSTPEEKLAALCKKYAELLLEHNSQKMLLQKQSQOLVOEKDHLGEGSKAVIARSK 251

QY 241 LSSICRELORHNSLKKEGVORAREEERKREKVTSHFOVTNDICLQEQHNSKLRQ 300

Db 252 LESICRELORHNSLKKEGVORAREEERKREKVTSHFOVTNDICLQEQHNSKLRQ 311

QY 301 ENMELAEELKLIQYELREBEHIDKVFHKLQOLQOLVDAKLQQAQEMLKBAERHOREKD 360

Db 312 ENMELAEELKLIQYELREBEHIDKVFHKLQOLQOLVDAKLQQAQEMLKBAERHOREKD 371

QY 361 FILKEAVESQRCMLKQOETHLQOOLALYTEKEEFPONTLSKSEVPTTFQEMEKMTK 420

Db 372 FILKEAVESQRCMLKQOETHLQOOLALYTEKEEFPONTLSKSEVPTTFQEMEKMTK 431

QY 421 KIKLEKETTMYRSWESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQT 476

Db 432 KIKLEKETTMYRSWESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQT 487

RESULT 6

AAW49042

ID AAW49042 standard; protein; 530 AA.

XX AAW49042;

AC AAW49042;

XX 09-NOV-1998 (first entry)

DT Human low density lipoprotein binding protein LBP-3.

DE Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;

KW receptor; human; atherosclerosis; diagnosis; therapy; vaccine.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 53..59

FT /note= "Claim 2"

XX WO9823282-A1.

XX 04-JUN-1998.

XX 26-NOV-1997; 97WO-US021857.

XX 27-NOV-1996; 96US-0031930P.

XX 03-JUN-1997; 97US-0048547P.

XX (BOST-) BOSTON HEART FOUND INC.

XX Lees AM, Lees RS, Law SW, Arjona AA;

XX WPI: 1998-322455/28.

DR N-PSDB; AAV32839.

XX Nucleic acid encoding low density lipoprotein binding proteins and related vectors - transformed cells, proteins, and modulators of binding, useful for treatment and diagnosis of atherosclerosis and for identifying subjects at risk.

XX Claim 1; Fig 8; 47pp; English.

XX This polypeptide comprises novel human low density lipoprotein (LDL) binding protein LBP-3 that is capable of binding both native and methyl

481 TDSGPERRPGQAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTPSARA 530

Db

RESULT 8

AB007223

ID ABO07223 standard; protein; 356 AA.

XX

AC

XX

DT

XX

DE

XX

KW

XX

KW

XX

KW

XX

KW

XX

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match

Best Local Similarity

Matches

65.2%; Score 356; DB 6; Length 356;

100.0%; Pred. No. 0;

Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

191

ALCKKYAEILLEHNRNSQKMKLLQKQSQQLVQKDHLEHNSKAVLARSKLESLCRLQK 250

DB

1

ALCKKYAEILLEHNRNSQKMKLLQKQSQQLVQKDHLEHNSKAVLARSKLESLCRLQK 60

QY

251

HNRSLKRGVQPARIEEKKREVTSHFQVTLNDIQLQMEQHNRNSKLEHNSKLESLCRLQK 310

DB

61

HNRSLKRGVQPARIEEKKREVTSHFQVTLNDIQLQMEQHNRNSKLEHNSKLESLCRLQK 120

QY

311

KLIEQVELREEHIDKVFHKKDQQLQVDAKQQAQEMLEKAEERHQRKDFLLKEAVESQ 370

DB

121

KLIEQVELREEHIDKVFHKKDQQLQVDAKQQAQEMLEKAEERHQRKDFLLKEAVESQ 180

QY

371

RMCCLMKQOETHLKOOLALYTEKFEFPONTLSKSSEVFTTFKQEVEMTKKIKLEKETT 430

DB

181

RMCCLMKQOETHLKOOLALYTEKFEFPONTLSKSSEVFTTFKQEVEMTKKIKLEKETT 240

QY

431

MYRSRWSSNKALLEWAEKTVRDKEGLQVKIQRLEKLCALQTERNDLNKRVQDLSA 490

DB

241

MYRSRWSSNKALLEWAEKTVRDKEGLQVKIQRLEKLCALQTERNDLNKRVQDLSA 300

QY

491

GGQGSILTDGSPERRPPEGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTPSARA 546

DB

301

GGQGSILTDGSPERRPPEGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTPSARA 356

RESULT 9

AAM78757

ID AAM78757 standard; protein; 358 AA.

XX

AC

AAM78757;

DT

06-NOV-2001 (first entry)

XX

DE

Human protein SEQ ID NO 1419.

XX

KW

Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW

vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW

tissue growth factor; immunomodulatory; cancer; leukaemia;

KW

nervous system disorder; arthritis; inflammation.

OS

Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AB007223

ID ABO07223 standard; protein; 356 AA.

XX

AC

XX

DT

13-AUG-2003 (first entry)

XX

DE

Human p53 modifying protein, SEQ ID 183.

XX

KW

Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;

KW

antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;

KW

lung cancer; ovarian cancer; angiogenesis; cell cycle;

KW

apoptotic disorder; cell proliferation disorder.

XX

OS

Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

The invention relates to identifying (M1) a candidate p53 pathway

modulating agent, by contacting an assay system comprising a purified HM

polypeptide (human orthologue of genes that modify the p53 pathway in

Drosophila) or nucleic acid with a test agent under conditions, where but

for the presence of the test agent, the system provides a reference

activity, and detecting a test agent-biased activity of the assay system.

Also included are modulating (M2) a p53 pathway of a cell (comprising

contacting a cell defective in p53 function with a candidate modulator

that specifically binds to a HM polypeptide comprising an HM amino acid

sequence, where p53 function is restored), modulating (M3) a p53 pathway

in a mammalian cell (comprising contacting the cell with an agent that

specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)

a disease in a patient (comprising: (a) obtaining a biological sample

from the patient; (b) contacting the sample with a probe for HM

expression; (c) comparing the results with a control; and (d) determining

whether the comparison indicates a likelihood disease). (M1) is useful

for identifying modulators of the p53 pathway. A probe for HM expression

is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,

in a patient, where the cancer has greater than 25 % expression level.

Modulators identified by (M1) are useful in a variety of diagnostic and

therapeutic applications, where disease or disorder prognosis is related

to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell

proliferation disorders (e.g. cancer). Another two new methods (M2 and

M3) are useful for modulating the p53 pathway of a cell, thus restoring

the p53 function of the cell, so that the cell undergoes normal

proliferation or progression through the cell cycle. (M2) and (M3) are

also useful for treating defects in the p53 pathway such as angiogenic,

apoptotic or cell proliferation disorders. The present sequence

represents a human p53 pathway modifying protein

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
PS Claim 20; Page 3681-3682; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 358 AA;

Query Match 63.6%; Score 347; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 LEEHNSOKMQLKQKQSQVLQVEKDHLEHSHKAVLARSKLSLCRELOHNSLKEEG 259
Db 12 LEEHNSOKMQLKQKQSQVLQVEKDHLEHSHKAVLARSKLSLCRELOHNSLKEEG 71

Qy 260 VQARAESEKKEVTSHTFQVTLNDIOLQMEQHNSKLRQENNELAERLKGLIEQVELR 319
Db 72 VQARAESEKKEVTSHTFQVTLNDIOLQMEQHNSKLRQENNELAERLKGLIEQVELR 131

Qy 320 EEHIDKVFPHKDLQQLVDALQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELMKQ 379
Db 132 EEHIDKVFPHKDLQQLVDALQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELMKQ 191

Qy 380 ETLKQQLALYTERKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRWS 439
Db 192 ETLKQQLALYTERKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRWS 251

Qy 440 NKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQGLTDS 499
Db 252 NKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQGLTDS 311

Qy 500 GPERPFGCAQAPSSPRVTEAPCYGAPSTEASGQTGPQPTPSARA 546
Db 312 GPERPFGCAQAPSSPRVTEAPCYGAPSTEASGQTGPQPTPSARA 358

RESULT 10
ABG00839
ID ABG00839 standard; protein; 1749 AA.
XX
AC ABG00839;
XX
XX 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #830.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR
PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
PA Dmanac RT, Liu C, Tang YT;
XX
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS65026.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 31198; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1749 AA;

Query Match 41.8%; Score 228; DB 4; Length 1749;
Best Local Similarity 100.0%; Pred. No. 1.1e-211;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 257 EGVQARAESEKKEVTSHTFQVTLNDIOLQMEQHNSKLRQENNELAERLKGLIEQY 316
Db 438 EGVQARAESEKKEVTSHTFQVTLNDIOLQMEQHNSKLRQENNELAERLKGLIEQY 497

Qy 317 ELREEHIDKVFPHKDLQQLVDALQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 376
Db 498 ELREEHIDKVFPHKDLQQLVDALQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 557

Qy 377 KQOETHLKQQLALYTERKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 436
Db 558 KQOETHLKQQLALYTERKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 617

Qy 437 ESSNKALLERMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNK 484
Db 618 ESSNKALLERMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNK 665

RESULT 11
AAW49039
ID AAW49039 standard; protein; 557 AA.
XX
AC AAW49039;
XX
XX 09-NOV-1998 (first entry)
XX
XX Rabbit low density lipoprotein binding protein LBP-3.
XX
XX Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
XX receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.
XX

OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT Peptide 96..110
FT /note="Claim 2"
XX
XX WO9823282-A1.
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-US021857.
XX
XX 27-NOV-1996; 96US-0031930P.
XX
XX 03-JUN-1997; 97US-0048547P.
XX
XX (BOST-) BOSTON HEART FOUND INC.
XX
XX Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 1998-322455/28.
XX
XX N-PSDB; AAV32836.
XX
XX Nucleic acid encoding low density lipoprotein binding proteins and
XX related vectors - transformed cells, proteins, and modulators of binding,
XX useful for treatment and diagnosis of atherosclerosis and for identifying
XX subjects at risk.
XX
XX Claim 1; Fig 5; 47pp; English.
XX
XX This polypeptide comprises novel rabbit low density lipoprotein (LDL)
XX binding protein LBP-3 that is capable of binding both native and methyl
XX LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA
XX (see AAV32836). cDNA clones (see AAV32834-39) and/or rabbit and human
XX LBP3 (see AAV49037-42) are claimed. An abnormality in an aspect of LBP
XX metabolism or structure is diagnostic of a risk for atherosclerosis. The
XX invention provides: methods for determining if an animal is at risk for
XX atherosclerosis (e.g. for prenatal screening); methods for treating
XX atherosclerosis (including gene therapy) using e.g. LBP polypeptides to
XX bind LDL and thereby prevent formation of atherosclerotic plaque; and
XX methods for treating a cell having an abnormality in LBP structure or
XX metabolism. Pharmaceutical and vaccine compositions are also provided, as
XX well as recombinant vectors and host cells used to produce recombinant
XX LBP
XX
XX Sequence 557 AA;
XX
XX Query Match 31.5%; Score 172; DB 2; Length 557;
XX Best Local Similarity 100.0%; Pred. No. 9.7e-158;
XX Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 280 TLNDIQLQMEQHNERNSKLRQENMELAEKLLKLIQYELREHIDKVFHKDLQQLVDA 339
XX
XX Db 280 TLNDIQLQMEQHNERNSKLRQENMELAEKLLKLIQYELREHIDKVFHKDLQQLVDA 339
XX
XX QY 340 KLQQAQEMLKEABERHOREKDFLLKAVESQRCMELMKQETHLKOQALALYTEKPEEFQ 399
XX
XX Db 340 KLQQAQEMLKEABERHOREKDFLLKAVESQRCMELMKQETHLKOQALALYTEKPEEFQ 399
XX
XX QY 400 TLSKSSSEVFTTFKQEMKMTKKIKLEKETTYRSRWESSNKALLEMAEKT 451
XX
XX Db 400 TLSKSSSEVFTTFKQEMKMTKKIKLEKETTYRSRWESSNKALLEMAEKT 451
XX
XX RESULT 12
XX AAB82801
XX ID AAB82801 standard; protein; 557 AA.
XX
XX AC AAB82801;
XX
XX XX 12-NOV-2001 (first entry)
XX
XX DE Rabbit low density lipoprotein binding protein 3 (LBP-3).
XX
XX KW Human breast cancer associated antigen protein sequence SEQ ID NO:621.
XX
XX KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX cancer associated antigen; cytostatic; cancer vaccine.
XX

KW Low density lipoprotein binding protein 3; LBP-3; LDL; rabbit;
KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX
OS Oryctolagus cuniculus.
XX
XX WO200164874-A2.
XX
XX 07-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-US006356.
XX
XX 02-MAR-2000; 2000US-00517849.
XX
XX 14-JUL-2000; 2000US-00616289.
XX
XX (BOST-) BOSTON HEART FOUND INC.
XX
XX Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 2001-565505/63.
XX
XX N-PSDB; AAV26491.
XX
XX New isolated low density lipoprotein binding polypeptide for treating,
XX diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX
XX Claim 13(e); Fig 5; 143pp; English.
XX
XX The present sequence is that of novel rabbit low density lipoprotein
XX binding protein 1 (LBP-3). The amino acid sequence is deduced from an
XX isolated cDNA clone (see AAV26491). Rabbit LBP-3 is an example of claimed
XX polypeptides of the invention, termed LBP3, that are capable of binding
XX to native and methylated low density lipoproteins. Also claimed are
XX biologically active fragments and analogues of LBP3, polynucleotides
XX encoding LBP3, as well as expression vectors, cells and methods of
XX producing the LBP3. A polypeptide having the amino acid residues 96-110
XX of the present sequence is claimed (see AAB82819). Methods of determining
XX if an animal is at risk for atherosclerosis, methods for evaluating an
XX agent for use in treating atherosclerosis, and methods for treating a
XX cell having an abnormality in structure or metabolism of LBP are claimed.
XX Pharmaceutical compositions comprising an LBP polypeptide or nucleic
XX acid, and vaccine compositions, are also claimed
XX
XX Sequence 557 AA;
XX
XX Query Match 31.5%; Score 172; DB 4; Length 557;
XX Best Local Similarity 100.0%; Pred. No. 9.7e-158;
XX Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 280 TLNDIQLQMEQHNERNSKLRQENMELAEKLLKLIQYELREHIDKVFHKDLQQLVDA 339
XX
XX Db 280 TLNDIQLQMEQHNERNSKLRQENMELAEKLLKLIQYELREHIDKVFHKDLQQLVDA 339
XX
XX QY 340 KLQQAQEMLKEABERHOREKDFLLKAVESQRCMELMKQETHLKOQALALYTEKPEEFQ 399
XX
XX Db 340 KLQQAQEMLKEABERHOREKDFLLKAVESQRCMELMKQETHLKOQALALYTEKPEEFQ 399
XX
XX QY 400 TLSKSSSEVFTTFKQEMKMTKKIKLEKETTYRSRWESSNKALLEMAEKT 451
XX
XX Db 400 TLSKSSSEVFTTFKQEMKMTKKIKLEKETTYRSRWESSNKALLEMAEKT 451
XX
XX RESULT 13
XX AAB63259
XX ID AAB63259 standard; protein; 204 AA.
XX
XX AC AAB63259;
XX
XX XX 26-MAR-2001 (first entry)
XX
XX DE Human breast cancer associated antigen protein sequence SEQ ID NO:621.
XX
XX KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX cancer associated antigen; cytostatic; cancer vaccine.
XX

```

OS Homo sapiens.
XX WO2000073801-A2.
XX PD 07-DEC-2000.
XX PF 26-MAY-2000; 2000WO-US014749.
XX PR 28-MAY-1999; 99US-0136526P.
XX PR 10-SEP-1999; 99US-0153454P.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX FI Obata Y;
XX DR WP1; 2001-025274/03.
XX PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer.
XX PS Example 1; Page 486-487; 799pp; English.
XX CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX represent nucleotide sequences encoding human breast, gastric and
XX prostate cancer associated antigen precursors (CAAP), respectively.
XX CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
XX represent human breast, gastric and prostate CAAP protein sequence
XX respectively. CAAPs have cytostatic activity and can be used in the
XX production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX condition characterised by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX SQ Sequence 204 AA;

Query Match 3.5%; Score 19; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEEKLAALCKKYA 197
Db 1 LNTLSTPEEKLAALCKKYA 19
|||||
|||||

RESULT 14
AAW37883
XX ID AAW37883 standard; protein; 386 AA.
XX AC AAW37883;
XX DT 28-AUG-1998 (first entry)
XX DE BRCA1 modulator protein 091-132Q20.
XX KW BRCA1 modulator protein; 091-132Q20; breast cancer antigen 1;
XX KW tumour suppressor protein; diagnosis; therapy; human.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 124..143
XX FT /note= "leucine zipper motif"
XX FN WO9810066-A1.
XX PD 12-MAR-1998.
XX PF 06-AUG-1997; 97WO-US013944.
XX PR 04-SEP-1996; 96US-0025601P.
XX PT

(ONYX-) ONYX PHARM INC.
Rubinfeld B, Polakis P, Ligenfelter C, Vuong TT;
WPI; 1998-193616/17.
N-PSDB; AAV29064.
Breast cancer antigen 1 modulator protein - useful for diagnosing
diseases involving unwanted cell growth, e.g. breast cancer, and for
producing therapeutics for treatment of such diseases.
Example 1; Fig 3; 73pp; English.
This polypeptide comprises a 46 kDa BRCA1 modulator protein that binds to
the tumour suppressor gene product BRCA1, and which is characterised by a
leucine zipper motif. Its amino acid sequence was deduced from the
nucleotide sequence of a cDNA clone (see AAV29064), designated 091-132Q20
(ATCC 98143), isolated from a HeLa cell cDNA library using a yeast two-
hybrid assay. 3 cDNA clones (see also AAV29062 and AAV29063) coding for
BRCA1 modulator proteins (see AAW37881-83) have been characterised.
Vectors and host cells comprising the isolated nucleic acid sequences are
claimed, as well as a process for producing BRCA1 modulator protein by
culturing these host cells. BRCA1 modulator proteins and nucleic acids
can be used to diagnose diseases involving unwanted cell growth, e.g.
breast cancer, and to identify compounds that alter BRCA1 interaction
with BRCA1 modulators for the treatment of such diseases
XX SQ Sequence 386 AA;

Query Match 3.5%; Score 19; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEEKLAALCKKYA 197
Db 135 LNTLSTPEEKLAALCKKYA 153
|||||
|||||

RESULT 15
AAV30151
XX ID AAV30151 standard; protein; 386 AA.
XX AC AAV30151;
XX DT 27-OCT-1999 (first entry)
XX DE Amino acid sequence of a BRCA1 modulator protein.
XX KW Modulator protein; BRCA1; tumour suppressor protein; breast cancer;
XX KW ovarian cancer; cell growth; cell proliferation.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 125..143
XX FT /note= "leucine zipper motif"
XX FN US5948643-A.
XX PD 07-SEP-1999.
XX PF 13-AUG-1997; 97US-00968751.
XX PR 13-AUG-1997; 97US-00968751.
XX PA (ONYX-) ONYX PHARM INC.
XX PI Rubinfeld B, Ligenfelter C, Vuong TT, Polakis P;
XX WPI; 1999-517952/43.
XX DR N-PSDB; AAX86756.
XX PT Modulator proteins that bind to and modulate the activity of the BRCA1

```


OM protein - protein search, using sw model

Run on: June 8, 2004, 16:36:07 ; Search time 21 seconds
(without alignments)
2500.980 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 546

Sequence: 1 MKNQDKKGAAGKSNFKSP.....APSTEASGQTGPQPTSPARA 546

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	17	3.1	676	2	JC7222	77K muscle-derived
2	9	1.6	594	2	S50611	hypothetical prote
3	9	1.6	1402	2	I46707	translation initia
4	8	1.5	162	2	T29728	hypothetical prote
5	8	1.5	211	2	T32354	hypothetical prote
6	8	1.5	224	2	T52558	translation elonga
7	8	1.5	229	2	E87236	probable membrane
8	8	1.5	260	2	S71315	deoxyguanosine kin
9	8	1.5	281	2	C88638	protein P58f6.1 [i
10	8	1.5	312	2	T08985	hypothetical prote
11	8	1.5	450	2	E71909	hypothetical prote
12	8	1.5	467	2	D88710	protein C43G2.1 [i
13	8	1.5	500	2	H97769	hypothetical prote
14	8	1.5	525	2	H75514	glutamyl-trNA synt
15	8	1.5	531	2	A55877	caldesmon, non-mus
16	8	1.5	535	2	C36811	hypothetical prote
17	8	1.5	535	2	A37994	RPI protein - balm
18	8	1.5	568	2	E30364	hypothetical prote
19	8	1.5	594	2	S62141	transcription init
20	8	1.5	758	2	S65169	hypothetical prote
21	8	1.5	761	2	T32193	hypothetical prote
22	8	1.5	928	2	T52232	endopeptidase Clp
23	8	1.5	952	2	T52456	endopeptidase Clp
24	8	1.5	952	2	T49283	AtClpC - Arabidops
25	8	1.5	1100	2	AE3243	conjugal transfer
26	8	1.5	1101	2	T03419	traA protein - Agr
27	8	1.5	1902	2	C97702	cell surface anti
28	8	1.5	2052	2	T18519	myosin X - bovine
29	7	1.3	23	2	S45032	homeotic protein S

hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
ribosomal protein
cystatin C - rat
conserved hypothet
cystatin C precurs
hypothetical prote
major allergen Par
hypothetical prote
hypothetical prote
major allergen Par
conserved hypothet
hypothetical prote
hypothetical prote

ALIGNMENTS

RESULT 1

JC7222
77K muscle-derived protein - chicken
C:Species: Gallus gallus (chicken)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C/Accession: JC7222
R/Dyeda, A.; Fukui, I.; Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T
Biochem. Biophys. Res. Commun. 269, 564-569, 2000
A/Title: MDP77: A novel neurite-outgrowth-promoting protein predominantly expressed in
A/Reference number: JC7222; MUID:20175243; PMID:10708594
A/Accession: JC7222
A/Molecule type: mRNA
A/Residues: 1-676 <UYE>
A/Cross-references: GB:D89999; NID:G7619883; PIDN:BAA94755.1; PID:G7619884
A/Experimental source: crus muscle
C/Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is i
C/Keywords: coiled coil, glycoprotein; leucine zipper; muscle

Query Match 3.1%; Score 17; DB 2; Length 676;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 LARSKLESICRLQRHN 252
DB 207 LARSKLESICRLQRHN 223

RESULT 2

S50611
hypothetical protein YER108c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C/Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 23-Mar-2001
C/Accession: S50611
R/Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A/Description: The sequence of S. cerevisiae cosmids 9781, 8198, 9115, 9981, and lambda
A/Reference number: S50437
A/Accession: S50611
A/Molecule type: DNA
A/Residues: 1-594 <PIE>
A/Cross-references: EMBL:U18916; NID:G1384128; PIDN:AC03206.1; PID:G603347; MIPS:YER10
C/Genetics:
A/Accession: S50611
A/Map position: 5R

Query Match 1.6%; Score 9; DB 2; Length 594;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 NTLSTPEEK 188

Db 466 NTLSTPREK 474

RESULT 3

I46707
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 31-Mar-2000
 C:Accession: I46707
 R:Yan, R.; He, W.; Rhoads, R.E.
 J. Biol. Chem. 268, 19200-19203, 1993
 A:Title: Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of
 A:Reference number: I46707; MUID:93374895; PMID:8396129
 A:Accession: I46707
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1402 <YAN>
 A:Cross-references: GB:L22090; MID:g404774; PIDN:AAA31242.1; PID:g404775

Query Match 1.6%; Score 9; DB 2; Length 1402;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 SEELSRQLE 80
 |||||
 Db 1243 SEELSRQLE 1251

RESULT 4

T29728
 hypothetical protein K03B4.7 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29728
 R:Du, Z.; Le, T.T.; Kemp, K.
 submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans cosmid K03B4.
 A:Reference number: Z20673
 A:Accession: T29728
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-162 <DUZ>
 A:Cross-references: EMBL:U55370; PIDN:AAA97999.1; GSPDB:GN00023; CESP:K03B4.7
 A:Experimental source: strain Bristol N2; clone K03B4
 C:Genetics:
 A:Gene: CESP:K03B4.7
 A:Map position: 5
 A:Introns: 58/3; 105/3

Query Match 1.5%; Score 8; DB 2; Length 162;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AERLKKLI 313
 |||||
 Db 89 AERLKKLI 96

RESULT 5

T32354
 hypothetical protein C08E3.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32354
 R:Miller, N.; Kramer, J.; Kepler, D.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid C08E3.
 A:Reference number: Z21155
 A:Accession: T32354
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-211 <MIL>
 A:Cross-references: EMBL:AF025457; PIDN:AA870965.1; GSPDB:GN00020; CESP:C08E3.4

A:Experimental source: strain Bristol N2; clone C08E3

C:Genetics:
 A:Gene: CESP:C08E3.4
 A:Map position: 2
 A:Introns: 36/3; 149/3

Query Match 1.5%; Score 8; DB 2; Length 211;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AERLKKLI 313
 |||||
 Db 81 AERLKKLI 88

RESULT 6

T52558
 translation elongation factor eEF1Balpha (clone 2) [validated] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 07-Jul-2003
 C:Accession: T52558
 R:Hericourt, P.; Jupin, I.
 FEBS Lett. 464, 148-152, 1999
 A:Title: Molecular cloning and characterization of the Arabidopsis thaliana alpha-subu
 A:Reference number: Z26114
 A:Accession: T52558
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-224 <HER>
 A:Cross-references: EMBL:AJ249597; PIDN:CAB64730.1

C:Genetics:
 A:Gene: eEF1Balpha2
 C:Function:
 A:Description: involved in translation elongation; able to complement a mutant yeast s

C:Superfamily: translation elongation factor eEF-1 beta chain; translation elongation f

Query Match 1.5%; Score 8; DB 2; Length 224;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 AAPAVEAE 44
 |||||
 Db 84 AAPAVEAE 91

RESULT 7

E87236
 probable membrane protein. [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: E87236
 R:Coles, S.T.; Eiglmeier, K.; Parthill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; F
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: E87236
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-229 <SRO>
 A:Cross-references: GB:AL450380; MID:g13093841; PIDN:CAC32147.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: ML2615
 C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0200

Query Match 1.5%; Score 8; DB 2; Length 229;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 PSQAAPAV 41
 |||||

Db	200	PSQAPAV	207	Db	241	GGPGEDGA	248
RESULT 8				RESULT 10			
S71315				T08985			
deoxyguanosine kinase (EC 2.7.1.113) precursor - human				hypotheical protein F6G3.120 - Arabidopsis thaliana			
C:Species: Homo sapiens (man)				C:Species: Arabidopsis thaliana (mouse-ear cress)			
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000				C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999			
C:Accession: S71315; S78432				C:Accession: T08985			
R:Wang, L.; Hellman, U.; Eriksson, S.				R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.P.			
FEBS Lett. 390, 39-43, 1996				submitted to the Protein Sequence Database, May 1999			
A:Title: Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA.				A:Reference number: Z16520			
A:Reference number: S71315; MUID:96314545; PMID:8706825				A:Accession: T08985			
A:Accession: S71315				A:Residues: 1-312 <BEV>			
A:Molecule type: mRNA				A:Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.120			
A:Residues: 1-260 <WAW>				A:Experimental source: cultivar Columbia; BAC clone F6G3			
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1				C:Genetics:			
A:Experimental source: tissue brain				A:Gene: ATSP:F6G3.120			
R:Wang, L.; Hellman, U.; Eriksson, S.				A:Map position: 4			
submitted to the EMBL Data Library, April 1996				A:Introns: 55/1; 94/3; 115/3; 151/3; 168/3; 189/3; 241/3; 264/3; 285/2			
A:Description: Cloning and expression of human deoxyguanosine kinase cDNA.				Query Match 1.5%; Score 8; DB 2; Length 312;			
A:Reference number: S78432				Best Local Similarity 100.0%; Pred. No. 13; Mismatches 0; Indels 0; Gaps 0;			
A:Accession: S78432				Matches 8; Conservative 0;			
A:Molecule type: mRNA				QY 342 QQAQEMLK 349			
A:Residues: 1-18; R', 20-260 <WAW>				DB 240 QQAQEMLK 247			
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1				RESULT 11			
A:Experimental source: brain				E71909			
C:Genetics:				hypotheical protein jhp0612 - Helicobacter pylori (strain J99)			
A:Gene: dgk				C:Species: Helicobacter pylori			
A:Genome: nuclear				A:Variety: strain J99			
C:Superfamily: human deoxycytidine kinase				C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999			
C:Keywords: mitochondrion; phosphotransferase				R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.			
F:1-22/Domain: transit peptide (mitochondrion) #status predicted <TNP>				; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.B.; Vovis, G.F.;			
F:23-260/Product: deoxyguanosine kinase #status predicted <MAT>				Nature 397, 176-180, 1999			
F:125-132/Region: DRS mctif				A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat			
F:185-191/Region: arginine-rich				A:Reference number: A71800; MUID:99120557; PMID:9923682			
Query Match 1.5%; Score 8; DB 2; Length 260;				A:Accession: E71909			
Best Local Similarity 100.0%; Pred. No. 11;				A:Status: preliminary			
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				A:Molecule type: DNA			
QY 261 QRAREEE 268				A:Residues: 1-450 <ARN>			
DB 188 QRAREEE 195				A:Cross-references: GB:AB001493; GB:AB001439; NID:94155161; PIDN:AAD06193.1; PID:941551			
RESULT 9				A:Experimental source: strain J99			
C88638				C:Genetics:			
protein F58F6.1 [imported] - Caenorhabditis elegans				A:Gene: jhp0612			
C:Species: Caenorhabditis elegans				Query Match 1.5%; Score 8; DB 2; Length 450;			
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Sep-2003				Best Local Similarity 100.0%; Pred. No. 18;			
C:Accession: C88638				Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
R:anonymous, The C. elegans Sequencing Consortium.				QY 79 LEDILSTY 86			
Science 282, 2012-2018, 1998				DB 283 LEDILSTY 290			
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog				RESULT 12			
A:Reference number: A75000; MUID:99069613; PMID:9851916				D88710			
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele				protein C43G2.1 [imported] - Caenorhabditis elegans			
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and				C:Species: Caenorhabditis elegans			
A:Accession: C88638				C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001			
A:Status: preliminary				C:Accession: D88710			
A:Molecule type: DNA				R:anonymous, The C. elegans Sequencing Consortium.			
A:Residues: 1-281 <STO>				Science 282, 2012-2018, 1998			
A:Cross-references: GB:chr_IV; PIDN:AAB88358.1; PID:92662599; GSPDB:GN00022; CBSP:F58F6.				A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo			
C:Genetics:				A:Reference number: A75000; MUID:99069613; PMID:9851916			
A:Gene: F58F6.1				A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele			
A:Map position: 4				A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and			
Query Match 1.5%; Score 8; DB 2; Length 281;				A:Accession: D88638			
Best Local Similarity 100.0%; Pred. No. 12;				A:Molecule type: DNA			
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				A:Residues: 1-281 <STO>			
QY 93 GGPGEDGA 100				A:Cross-references: GB:chr_IV; PIDN:AAB88358.1; PID:92662599; GSPDB:GN00022; CBSP:F58F6.			
				C:Genetics:			
				A:Gene: F58F6.1			
				A:Map position: 4			
				Query Match 1.5%; Score 8; DB 2; Length 281;			
				Best Local Similarity 100.0%; Pred. No. 12;			
				Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
				QY 93 GGPGEDGA 100			

A:Accession: D88710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-467 <STO>
A:Cross-references: GB:chr_IV; PIDN:AA09107.1; PID:gl572756; GSPDB:GN00022; CBSP:C43G2
C:Genetics:
A:Gene: C43G2.1
A:Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 LAALCKXY 136
|||
Db 22 LAALCKXY 29

RESULT 13

H97769
Hypothetical protein murD [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: H97769
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rickettsia conorii
Science 293, 2093-2096, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: H97769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-500 <KUR>
A:Cross-references: GB:AR006914; PIDN:AAU03098.1; PID:gl5619640; GSPDB:GN00173
C:Genetics:
A:Gene: murD
C:Superfamily: UDP-N-acetylmuramate-alanine ligase

Query Match 1.5%; Score 8; DB 2; Length 500;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 QGEPARPE 108
|||
Db 278 QGEPARPE 285

RESULT 14

H75514
glutamy1-trNA synthetase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75514
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.N.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-525 <WHI>
A:Cross-references: GB:AR001907; GB:AE000513; NID:96458162; PIDN:AAF10063.1; PID:9645817
C:Experimental source: strain R1
C:Genetics:
A:Gene: DR0485
A:Map position: 1
C:Superfamily: glutamate-trNA ligase; glutamine-trNA ligase homology

Query Match 1.5%; Score 8; DB 2; Length 525;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 512 APSSPRVT 519
|||
Db 45 APSSPRVT 52

RESULT 15

A55887
caldesmon, non-muscle - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999
C:Accession: A55887
R:Yamashiro, S.; Yamakita, Y.; Yoshida, K.; Takiguchi, K.; Matsumura, F.
J. Biol. Chem. 270, 4023-4030, 1995
A:Title: Characterization of the COOH terminus of non-muscle caldesmon mutants lacking
A:Reference number: A55887; MUID:95181370; PMID:7876150
A:Accession: A55887
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-531 <YAM>
A:Cross-references: GB:U18419; NID:9622966; PIDN:AAA68521.1; PID:9622967
A:Note: authors translated the codon GCC for residue 68 as Val
C:Superfamily: caldesmon
C:Keywords: phosphoprotein

Query Match 1.5%; Score 8; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 AREEREKR 270
|||
Db 334 AREEREKR 341

Search completed: June 8, 2004, 16:39:54
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:34:26 ; Search time 17 Seconds
(without alignments)
1672.371 Million cell updates/sec

Title: US-10-023-529-44
Perfect score: 546
Sequence: 1 MKNQDKNGAKQSNPKSP.....APSTEASGQTGPQEPSTAPA 546

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259	47.4	259	YL14 HUMAN	P40222 homo sapien
2	9	1.6	799	FLO8 YEAST	Q8068 saccharomyc
3	9	1.6	1070	CARB METAC	Q8068 methanosarc
4	9	1.6	1073	CARB METWA	P58944 methanosarc
5	9	1.6	1402	IP4G RABIT	P41110 oryctolagus
6	8	1.5	483	SYE DEIRA	Q9rx30 deinococcus
7	8	1.5	500	MURD RICON	Q92i60 rickettsia
8	8	1.5	531	CALD RAT	Q82736 rattus norv
9	8	1.5	535	VG50 HSVSA	Q01012 herpesvirus
10	8	1.5	594	TFCS YEAST	P46678 saccharomyc
11	8	1.5	758	YPS8 YEAST	Q99299 saccharomyc
12	8	1.5	1070	RPOB LOTJA	Q9bb99 lotus japon
13	8	1.5	1072	CARB THETN	Q8rbx0 thermoanaer
14	8	1.5	1100	TPAA AGRTS	Q44349 agrobacteri
15	8	1.5	2052	MT10 EOVIN	P79114 bos taurus
16	7	1.3	105	RK21 ODOSI	P49557 oodentelia s
17	7	1.3	121	INSC ECOLI	P19776 escherichia
18	7	1.3	121	INSC SHIFL	P59444 shigella fl
19	7	1.3	127	CYTC RAT	P14841 rattus norv
20	7	1.3	138	NL13 PARHU	Q40905 parietaria
21	7	1.3	139	NL11 PARJU	P43217 parietaria
22	7	1.3	176	NL12 PARJU	Q04404 parietaria
23	7	1.3	186	RRP WOLBU	Q7mad8 wolfinella s
24	7	1.3	187	ISP2 VIBPA	Q87na5 vibrio para
25	7	1.3	187	ISP2 VIBU	P59366 vibrio vuln
26	7	1.3	193	LOLB NEIMA	P57023 neisseria m
27	7	1.3	193	LOLB NEIMS	P57024 neisseria m
28	7	1.3	195	AAAT HDVD3	P39995 hepatitis d
29	7	1.3	195	AAAT HDVU1	P39833 hepatitis d
30	7	1.3	195	AAAT HDVM1	P5881 hepatitis d
31	7	1.3	195	AAAT HDVM2	P25882 hepatitis d
32	7	1.3	199	VATE PYRAB	Q9uxu4 pyrococcus
33	7	1.3	207	TRPF_STAP	Q8csn5 staphylococ

34	7	1.3	211	1	YA03 ARCFU	O29259 archaeoglob
35	7	1.3	214	1	AAANT HDVAM	P25989 hepatitis d
36	7	1.3	214	1	AAANT HDVNA	P25880 hepatitis d
37	7	1.3	218	1	SGAH MYCPN	P75293 mycoplasma
38	7	1.3	220	1	R929 HUMAN	O95707 homo sapien
39	7	1.3	221	1	YNOC_ECOLI	P76114 escherichia
40	7	1.3	225	1	TIUY ARATH	Q98w33 arabidopsis
41	7	1.3	227	1	PYRH AERPE	Q9yfc4 aeropyrum p
42	7	1.3	229	1	REGQ BP82	P13870 bacterioph
43	7	1.3	248	1	MYFO HUMAN	P25189 homo sapien
44	7	1.3	277	1	DGK HUMAN	Q16854 homo sapien
45	7	1.3	297	1	PERE_ECOLI	Q57083 escherichia

ALIGNMENTS

RESULT 1
YL14 HUMAN
ID YL14 HUMAN STANDARD; PRT; 259 AA.
AC P40222; Q8N2Y3;
DT 01-FEB-1995 (Rel. 31, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein initially thought to be identical with
interleukin-14 (IL-14) (High molecular weight B-cell growth factor)
DE (HMW-BCGP).
DE IL14.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93317675; PubMed=8327514;
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RT "Identification of a cDNA for a human high-molecular-weight B-cell
growth factor."
RL Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).
RN [2]
RP ERRATUM.
RX MEDLINE=96353961; PubMed=8755619;
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RL Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).
RN [3]
RP SEQUENCE OF 130-259 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Pahay J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shcherbenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- CAUTION: The clone described in Ref.1 was initially thought to
code for a high molecular weight interleukin (IL-14). The
translation of this cDNA does not predict the open reading frame

CC for the 60 kDa protein described in Ref.1. A reading frame on the
CC plus strand predicts a 7.7 kDa protein. The longest open reading
CC frame (shown here) is on the opposite strand and predicts a 36.4
CC kDa protein. The relationship of this sequence to IL-14, if any,
CC is uncertain.

CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; L15344; -; NOT ANNOTATED CDS.
CC EMBL; BC029686; AAH25686.1; -;
CC PIR; A48203; A48203.
CC Genew; HGNC:5976; IL14.
CC MIM; 147684;
CC GO; GO:0005576; C:extracellular; NAS.
CC GO; GO:0003072; F:high molecular weight B-cell growth factor . . .; NAS.
CC GO; GO:0008283; P:cell proliferation; NAS.
CC KW Hypothetical protein.
CC SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;

CC -----
CC Query Match 47.4%; Score 259; DB 1; Length 259;
CC Best Local Similarity 100.0%; Pred. No. 1.1e-249; Indels 0; Gaps 0;
CC Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC -----
CC QY 288 MEQHNERSKLRQENMELAEKLEIYELNEEHIDKVFVKHDLQQQLVDAKLQQAQEM 347
CC Db 1 MEQHNERSKLRQENMELAEKLEIYELNEEHIDKVFVKHDLQQQLVDAKLQQAQEM 60
CC -----
CC QY 348 LKAEERHQREDFLLKKAESVQRCMLKQOETHLKOALYTKPEEFQNTLSKSEV 407
CC Db 61 LKAEERHQREDFLLKKAESVQRCMLKQOETHLKOALYTKPEEFQNTLSKSEV 120
CC -----
CC QY 408 FTTTQKQEMKTKKIKLEKETTYRSWESSKALLEAEKTVRKELSGLVKIQRL 467
CC Db 121 FTTTQKQEMKTKKIKLEKETTYRSWESSKALLEAEKTVRKELSGLVKIQRL 180
CC -----
CC QY 468 EKLCEALQTERNDLNKRVODLSAGQGGSLTDSGPPRRPFGGAQAPSSRVTEAPCTGA 527
CC Db 181 EKLCEALQTERNDLNKRVODLSAGQGGSLTDSGPPRRPFGGAQAPSSRVTEAPCTGA 240
CC -----
CC QY 528 PSTEASGQTGPQEPSARA 546
CC Db 241 PSTEASGQTGPQEPSARA 259

CC -----
CC RESULT 2
CC FLO8 YEAST STANDARD; PRT; 799 AA.
CC AC P40068; P40067; Q05751;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Transcriptional activator FLO8 (PDH5 protein).
CC GN FLO8 OR PDH5 OR YER109C/YER108C.
CC OS Saccharomyces cerevisiae (Baker's yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC SC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CC OX NCBI_TaxID=4932;
CC RN [1]
CC SEQUENCE FROM N.A.
CC RC STRAIN=Sigma 1278B;
CC RA Liu H., Styles C.A., Fink G.;
CC RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC [2]
CC SEQUENCE OF 71-799 FROM N.A.
CC RC STRAIN=ATCC 60715;
CC RX MEDLINE=96335146; PubMed=8757402;
CC RA Kobayashi O., Suda H., Ohtani T., Sone H.;

RT "Molecular cloning and analysis of the dominant flocculation gene
RT FLO8 from Saccharomyces cerevisiae.";
RT Mol. Gen. Genet. 251:707-715(1996).
RT [3]

RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=9733264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA On C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).

CC -!- FUNCTION: Required for diploid filamentous growth, haploid
CC invasive growth and flocculation. Putative transcriptional
CC activator of FLO1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift.

CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; U51431; AAC49522.1; -;
CC EMBL; D83713; BAA12076.1; -;
CC EMBL; U18916; AAC03207.1; ALT_FRAME.
CC EMBL; U18916; AAC03206.1; ALT_FRAME.
CC GenOnline; 139188; -;
CC TRANSFAC; T03311; -;
CC SGD; S0000911; FLO8.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0005634; C:nucleus; IDA.
CC InterPro; IPR006594; Lish.
CC SMART; SMO0667; Lish; 1.
CC PROSITE; PS00896; LISH; 1.
CC Transcription regulation; Activator; Nuclear protein.
CC DOMAIN 73 105 LISH.
CC FT DOMAIN 41 55 POLY-GLN.
CC FT CONFLICT 112 112 V -> I (IN REF. 2).
CC FT CONFLICT 115 115 P -> S (IN REF. 2).
CC FT CONFLICT 383 383 C -> G (IN REF. 2).
CC FT CONFLICT 441 441 A -> T (IN REF. 2).
CC FT CONFLICT 447 447 A -> V (IN REF. 2).
CC FT CONFLICT 598 598 R -> P (IN REF. 2).
CC SQ SEQUENCE 799 AA; 86834 MW; 91AAE10D8E586DAF CRC64;

CC -----
CC Query Match 1.6%; Score 9; DB 1; Length 799;
CC Best Local Similarity 100.0%; Pred. No. 1.8;
CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC -----
CC QY 180 NTLSTPEEK 188
CC Db 671 NTLSTPEEK 679

CC -----
CC RESULT 3
CC CARB METAC STANDARD; PRT; 1070 AA.
CC ID CARB METAC
CC AC Q8TNY4;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
CC phosphate synthetase ammonia chain).

GN CARB OR MA2143.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W.E., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.B., Kraybill J.A., Guss A.M.,
RA Hedgerich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayal L.A., White O., Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity".
RL Genome Res. 12:532-542(2002).
CC -!- CARBONIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010901; AM05541.1; --
CC HAMAP; MF_01210; 1;
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPhase_L.
CC InterPro; IPR005479; CPhase_L_D2.
CC InterPro; IPR005480; CPhase_L_D3.
CC InterPro; IPR005481; CPhase_L_N.
CC InterPro; IPR004362; MGS like.
CC Pfam; PF00289; CPhase_L_chain; 2.
CC Pfam; PF02786; CPhase_L_D2; 2.
CC Pfam; PF02787; CPhase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPSASE.
CC TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
CC PROSITE; PS00865; CPSASE_1; 1.
CC PROSITE; PS00867; CPSASE_2; 1.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
FT ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 399 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 400 540 OLIGOMERIZATION DOMAIN.
FT DOMAIN 541 931 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 932 1070 ALLOSTERIC DOMAIN.
FT REPEAT 1 540
FT REPEAT 541 1070
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 300 350 ATP (POTENTIAL).
FT METAL 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 296 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 822 MANGANESE 3 (BY SIMILARITY).
FT METAL 834 MANGANESE 3 (BY SIMILARITY).

SQ SEQUENCE 1070 AA; 118441 MW; 4C971B8C42B9C3B5 CRC64;
Query Match 1.6%; Score 9; DB 1; Length 1070;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 253 RSLKEGVQ 261
DB 35 RSLKEGVQ 43
RESULT 4
CARB_METWA STANDARD; PRT; 1073 AA.
ID CARB_METWA
AC P58944;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR MM0038.
OS Methanosarcina mazel (Methanosarcina frieia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goe1 / Goe1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierzer A., Baumeister S., Jacobi C.,
RA Bruggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazel: evidence for lateral gene
RT transfer between Bacteria and Archaea".
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE013223; AM29734.1; --
CC HAMAP; MF_01210; 1;
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPhase_L.
CC InterPro; IPR005479; CPhase_L_D2.
CC InterPro; IPR005480; CPhase_L_D3.
CC InterPro; IPR005481; CPhase_L_N.
CC InterPro; IPR004362; MGS like.
CC Pfam; PF00289; CPhase_L_chain; 2.
CC Pfam; PF02786; CPhase_L_D2; 2.
CC Pfam; PF02787; CPhase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPSASE.
CC TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
CC PROSITE; PS00865; CPSASE_1; 1.
CC PROSITE; PS00867; CPSASE_2; 1.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
FT ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 399 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 400 540 OLIGOMERIZATION DOMAIN.
FT DOMAIN 541 931 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 932 1070 ALLOSTERIC DOMAIN.
FT REPEAT 1 540
FT REPEAT 541 1070
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 300 350 ATP (POTENTIAL).
FT METAL 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 296 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 822 MANGANESE 3 (BY SIMILARITY).
FT METAL 834 MANGANESE 3 (BY SIMILARITY).

KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 399 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 400 540 OLIGOMERIZATION DOMAIN.
FT DOMAIN 541 931 CARDAMOMYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 932 1073 ALLOSTERIC DOMAIN.
FT REPEAT 1 540
FT REPEAT 541 1073
FT NP_BIND 153 210
FT NP_BIND 300 350
FT NP_BIND 284 284
FT METAL 296 296
FT METAL 298 298
FT METAL 822 822
FT METAL 834 834
SQ SEQUENCE 1073 AA; 118963 MW; ACBD9E5DFC1EAD1 CRC64;

Query Match 1.6%; Score 9; DB 1; Length 1073;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 RSLKEGVQ 261
DB 35 RSLKEGVQ 43

AC P41110; STANDARD; PRT; 1402 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Eukaryotic translation initiation factor 4 gamma (eIF-4-gamma) (eIF4G)
DE (eIF-4G) (p220).
GN EIP4G.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 479-500.
RC STRAIN=New Zealand white; TISSUE=Brain;
RX MEDLINE=93374895; PubMed=8396129;
RA Lamphear B.J., Fan X., Yang F., Waters D., Liebig H.-D.,
RA Klump H., Kuechler E., Skern T., Rhoads R.E.;
RT "Mapping the cleavage site in protein synthesis initiation factor
eIF-4 gamma of the 2A proteases from human Cocksackievirus and
rhinovirus."
RT J. Biol. Chem. 268:19200-19203(1993).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=93054654; PubMed=1429670;
RA Yan R., Rychlik W., Etchison D., Rhoads R.E.;
RT "Amino acid sequence of the human protein synthesis initiation factor
eIF-4 gamma."
RL J. Biol. Chem. 267:23226-23231(1992).
CC -!- FUNCTION: Component of the protein complex EIP4F, which is
involved in the recognition of the mRNA cap, ATP-dependent
unwinding of 5'-terminal secondary structure and recruitment of
mRNA to the ribosome.
CC -!- SUBUNIT: EIP4F is a multi-subunit complex, the composition of
which varies with external and internal environmental conditions.
CC It is composed of at least EIP4A, EIP4B and EIP4G. EIP4G interacts
CC with the serine/threonine kinases MNK1 and MNK2. Appears to act
CC as a scaffold protein, holding these enzymes in place to
CC phosphorylate EIP4E (By similarity).
CC -!- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL
CC REGION OF WHEAT EUKARYOTIC INITIATION FACTOR (ISO)4F SUBUNIT P82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L22090; AAA31242.1; -
DR PIR; I46707; 146707.
DR InterPro; IPR008938; ARM
DR InterPro; IPR003307; eIF5C
DR InterPro; IPR003890; IF_eIF4G_MA3.
DR InterPro; IPR003891; IF_eIF4G_MA3.
DR Pfam; PF02847; MA3; 1.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; PF02020; W2; 1.
DR SMART; SM00515; eIF5C; 1.
DR SMART; SM00544; MA3; 1.
DR SMART; SM00543; MIF4G; 1.
KW Initiation factor; Protein biosynthesis; Translation regulation;
KW Phosphorylation; RNA-binding.
FT DOMAIN 188 192 POLY-PRO.
FT DOMAIN 262 275 POLY-GLU.
FT DOMAIN 1393 1398 POLY-GLU.
SQ SEQUENCE 1402 AA; 154050 MW; 7FD85D7E30519230 CRC64;

Query Match 1.6%; Score 9; DB 1; Length 1402;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 SBELSQLE 80
DB 1243 SBELSQLE 1251

RESULT 6
ID_SVE DEIRA STANDARD; PRT; 483 AA.
AC Q9RX30;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (GluRS).
GN GUTX OR DR0485.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1."
RL Science 286:1571-1577(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC or send an email to license@isb-sib.ch).
CC -----
CC ENEL; AE008617; AAL03098.1; -.
CC DR PIR; H97769; H97769.
CC DR HAMAP; MF 00639; atypical; 1.
CC DR InterPro; IPR000713; Mur_ligase.
CC DR InterPro; IPR004101; Mur_ligase_C.
CC DR InterPro; IPR005782; MurD.
CC DR Pfam; PF01225; Mur_ligase; 1.
CC DR TIGRFAMs; TIGR01087; murD; 1.
CC DR TIGRFAMs; TIGR01087; murD; 1.
CC KW Peptidoglycan synthetase; Cell wall; Cell division; Ligase;
CC ATP-binding; Complete proteome.
CC FT NP_BIND 111 117 ATP (POTENTIAL).
CC FT DOMAIN 260 306 RPE3.
CC SQ SEQUENCE 500 AA; 55787 MW; 024B268F6074822C CRC64;

Query Match 1.5%; Score 8; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 APSSPRVT 519
DB 3 APSSPRVT 10

RESULT 7
ID MURD_RICCN STANDARD; PRT; 500 AA.
AC Q92160;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE UDP-N-acetylmuramoylalanine--D-glutamate ligase (EC 6.3.2.9) (UDP-N-
DE acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid
DE adding enzyme).
GN MURD OR RC0560.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the addition of glutamate
CC to the nucleotide precursor UDP-N-acetylmuramoyl-L-alanine (UMA
CC (by similarity)).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine +
CC glutamate = ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamate.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: Belongs to the murDEP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR ENEL; AE008617; AAL03098.1; -.
DR PIR; H97769; H97769.
DR HAMAP; MF 00639; atypical; 1.
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR InterPro; IPR005782; MurD.
DR Pfam; PF01225; Mur_ligase; 1.
DR TIGRFAMs; TIGR01087; murD; 1.
DR TIGRFAMs; TIGR01087; murD; 1.
KW Peptidoglycan synthetase; Cell wall; Cell division; Ligase;
KW ATP-binding; Complete proteome.
FT NP_BIND 111 117 ATP (POTENTIAL).
FT DOMAIN 260 306 RPE3.
SQ SEQUENCE 500 AA; 55787 MW; 024B268F6074822C CRC64;

Query Match 1.5%; Score 8; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 QGEPAPPE 108
DB 278 QGEPAPPE 285

RESULT 8
ID CALD_RAT STANDARD; PRT; 531 AA.
AC Q62736;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Non-muscle caldesmon (CDM) (l-caldesmon).
GN CALDI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., MUTAGENESIS, AND PHOSPHORYLATION SITES.
RC TISSUE=Liver;
RX MEDLINE=95181370; PubMed=7876150;
RA Yamashiro S., Yamakita Y., Yoshida K.-S., Takiguchi K., Matsumura F.;
RT "Characterization of the COOH terminus of non-muscle caldesmon
RT mutants lacking mitosis-specific phosphorylation sites.";
RN J. Biol. Chem. 270:4023-4030(1995).
RL [2]
RP PHOSPHORYLATION BY CDC2.
RX MEDLINE=91095023; PubMed=1986309;
RA Yamashiro S., Yamakita Y., Hosoya H., Matsumura F.;
RT "Phosphorylation of non-muscle caldesmon by p34cdc2 kinase during
RT mitosis.";
RL Nature 349:169-172(1991).
CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the
CC regulation of actomyosin interactions in smooth muscle and actin
CC filaments. Stimulates actin binding of tropomyosin which
CC increases the stabilization of actin filament structure. In muscle
CC tissues, inhibits the actomyosin ATPase by binding to P-actin.
CC This inhibition is attenuated by calcium-calmodulin and is
CC potentiated by tropomyosin. Interacts with actin, myosin, two
CC molecules of tropomyosin and with calmodulin. Also play an
CC essential role during cellular mitosis and receptor capping.
CC -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on
CC stress fibers in fibroblasts (nonmuscle) (By similarity).
CC -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon)
CC is predominantly expressed in smooth muscles, whereas low-
CC molecular-weight caldesmon (l-caldesmon) is widely distributed in
CC non-muscle tissues and cells. Not expressed in skeletal muscle or
CC heart (by similarity).
CC -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-
CC binding domain, and the C-terminal a tropomyosin/actin/calmodulin-
CC binding domain. These two domains are separated by a central
CC helical region in the smooth-muscle form.
```

```

CC CC -!- PTM: In non-muscle cells, phosphorylation by CDC2 during mitosis
CC CC causes caldesmon to dissociate from microfilaments.
CC CC Phosphorylation reduces caldesmon binding to actin, myosin, and
CC CC calmodulin as well as its inhibition of actomyosin ATPase
CC CC activity. Phosphorylation also occurs in both quiescent and
CC CC dividing smooth muscle cells with similar effects on the
CC CC interaction with actin and calmodulin and on microfilaments
CC CC reorganization.
CC CC
CC CC -!- SIMILARITY: Belongs to the caldesmon family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U18419; AAA68521.1; --
CC CC PIR; A55887; A55887;
CC CC InterPro; IPR006017; Caldesmon.
CC CC InterPro; IPR006018; Caldesmon_LSP.
CC CC Pfam; PF02029; Caldesmon; 1.
CC CC PRINTS; PR01076; CALDESMON.
CC CC Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
CC CC Alternative splicing.
CC CC
CC CC DOMAIN 20 200 MYOSIN AND CALMODULIN-BINDING (BY
CC CC SIMILARITY).
CC CC FT DOMAIN 303 360 TROPOMYOSIN-BINDING (POTENTIAL).
CC CC FT DOMAIN 402 412 TROPOMYOSIN-BINDING (POTENTIAL).
CC CC FT DOMAIN 392 424 STRONG ACTIN-BINDING (BY SIMILARITY).
CC CC FT DOMAIN 454 460 CALMODULIN-BINDING (BY SIMILARITY).
CC CC FT DOMAIN 506 531 WEAK ACTIN-BINDING (BY SIMILARITY).
CC CC FT DOMAIN 33 40 POLY-ARG.
CC CC FT DOMAIN 180 189 POLY-GLU.
CC CC FT DOMAIN 279 282 POLY-ARG.
CC CC FT DOMAIN 319 322 POLY-GLU.
CC CC FT DOMAIN 336 339 POLY-GLU.
CC CC FT MOD RES 249 249 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD RES 462 462 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD RES 468 468 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD RES 491 491 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD RES 497 497 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD RES 527 527 PHOSPHORYLATION (BY CDC2).
CC CC FT MUTAGEN 249 249 S->A: DECREASES STRONGLY PHOSPHORYLATION-
CC CC FT MUTAGEN 462 462 S->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 468 468 T->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 491 491 ACTIN BINDING.
CC CC FT MUTAGEN 497 497 S->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 497 497 S->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 527 527 S->A: DOES NOT DECREASE PHOSPHORYLATION-
CC CC FT MUTAGEN 527 527 ACTIN BINDING.
CC CC FT SEQUENCE 531 AA; 60584 MW; CBEC50271A23829 CRC64;
CC CC
CC CC Query Match 1.5%; Score 8; DB 1; Length 531;
CC CC Best Local Similarity 100.0%; Pred.No. 13;
CC CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 263 AREEEKR 270
CC CC DB 334 AREEEKR 341
CC CC
CC CC RESULT 9
CC CC ID_VG50_HSVSA STANDARD; PRT; 535 AA.
CC CC AC Q01012;
CC CC DT 01-APR-1993 (Rel. 25, Created)
CC CC DT 01-APR-1993 (Rel. 25, Last sequence update)

```

```

DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Probable transcription activator EDRF1.
GN 50 OR EDRF1.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92233688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA Nicholas J., Coles L.S., Newman C., Honess R.W.;
RT "Regulation of the herpesvirus saimiri (HVS) delayed-early
RT 110-kilodalton promoter by HVS immediate-early gene products and a
RT homolog of the Epstein-Barr virus R trans activator.";
RL J. Virol. 62:2457-2466(1988).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATION. REGULATES THE DELAYED-EARLY
CC 110 KDA PROMOTER.
CC -!- SIMILARITY: TO EBV BRLF1.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X64346; CA445672.1; ALT_INIT.
CC CC EMBL; M86409; AAA46124.1; ALT_INIT.
CC CC EMBL; M60850; AAA46159.1; ALT_INIT.
CC CC InterPro; IPR004998; Herpes_TAF50.
CC CC Pfam; PF03326; Herpes_TAF50; 1.
CC CC KW Transcription regulation; Activator; DNA-binding; Early protein.
CC CC SQ SEQUENCE 535 AA; 60050 MW; B4F2B9ABA38616FB CRC64;
CC CC
CC CC Query Match 1.5%; Score 8; DB 1; Length 535;
CC CC Best Local Similarity 100.0%; Pred.No. 13;
CC CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 167 GLGKEITL 174
CC CC DB 187 GLGKEITL 194
CC CC
CC CC RESULT 10
CC CC ID_TFC5_YEAST STANDARD; PRT; 594 AA.
CC CC AC P46678;
CC CC DT 01-NOV-1995 (Rel. 32, Created)
CC CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DS Transcription factor TFIIB B* component (TFIIB90).
CC CC TFC5 OR TFC7 OR YNL039W OR N2682.
CC CC GN Saccharomyces cerevisiae (Baker's yeast).
CC CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CC CC OX NCBI_TaxID=4932;
CC CC RN [1]

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=S288C / AB972;
RX Kassaravetis G.A., Nguyen S.T., Kobayashi R., Kumar A.,
RA Glodushchek E.P., Pisano M.;
RT "Cloning, expression, and function of TFC5, the gene encoding the B'
RT component of the Saccharomyces cerevisiae RNA polymerase III
transcription factor TFIIB";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9786-9790(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=9603864; PubMed=7568218;
RA Kassaravetis G.A., Nguyen S.T., Kobayashi R., Kumar A.,
RA Glodushchek E.P., Pisano M.;
RT "Cloning, expression, and function of TFC5, the gene encoding the B'
RT component of the Saccharomyces cerevisiae RNA polymerase III
transcription factor TFIIB";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9786-9790(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96203116; PubMed=8617241;
RA Ruetz J., Conesa C., Dieci G., Lefebvre O., Duesterhoeft A.,
RA Ottonello S., Sentenac A.;
RT "A suppressor of mutations in the class III transcription system
RT encodes a component of yeast TFIIB";
RL EMBO J. 15:1941-1949(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RA Roberts S., Miller S., Lane W.S., Hahn S.;
RA Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RC Duesterhoeft A., Floeth M., Fritz C., Reuss-Reitzel D., Hilbert H.,
RA Moesti D.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: General activator of RNA polymerase III transcription.
CC -I- SUBUNIT: TFIIB comprises the TATA-binding protein (TBP), the
CC B-related factor (BRF) and the B' component (TFC5).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: TO S.POMBE SPCC1919.14C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31819; AAC49073.1; -
DR EMBL; U38415; AAC49364.1; -
DR EMBL; U37533; AAC49348.1; -
DR EMBL; Z71315; CAA95906.1; -
DR PIR; S62141; S62141. -
DR GerMOnline; 143046; -
DR TRANSFAC; T00055; -
DR SGD; S0004984; TFC5.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
DR Transcription regulation; Activator; Nuclear protein.
KW Transcription regulation; Activator; Nuclear protein.
SQ SEQUENCE 594 AA; 67687 MW; 7B0115BBB2491175 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 594;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 EEEERKE 272
DB 334 EEEERKE 341
|||||

RESULT 11
YP58 YEAST STANDARD; PRT; 758 AA.
AC Q92959;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein YPL158C.

GN OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97103777; PubMed=8948103;
RA Furnelle B., Coster F., Goffeau A.;
RT "The sequence of 55 kb on the left arm of yeast chromosome XVI
RT identifies a small nuclear RNA, a new putative protein kinase and two
RT new putative regulators";
RL Yeast 12:1483-1492(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Neutwich U., Oefner P., Pearson D.,
RA Petal P.X., Pohl T.M., Furnelle B., Schafer M., Scharfe M.,
RA Schrems B., Schramm S., Schroeder M., Seicu A.M., Tettelin H.,
RA Urrastarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
RL Nature 387:103-105(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96770; CAA65563.1; -
DR EMBL; Z73514; CAA97863.1; -
DR PIR; S65169; S65169.
DR GerMOnline; 144140; -
DR SGD; S0006079; YPL158C.
DR Hypothetical protein; Coiled coil.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 253 263 POLY-SER.
FT DOMAIN 648 711 COILED COIL (POTENTIAL).
FT DOMAIN 652 658 POLY-GLU.
FT DOMAIN 664 669 POLY-GLU.
FT DOMAIN 686 690 POLY-GLU.
SQ SEQUENCE 758 AA; 84845 MW; 3B4FA92B91C87F2B CRC64;

Query Match 1.5%; Score 8; DB 1; Length 758;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 EEEERKE 272
DB 687 EEEERKE 694
|||||

RESULT 12
RPOB LOTJA STANDARD; PRT; 1070 AA.
ID RPOB LOTJA
AC Q9B59;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)
DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
RPOB
Lotus japonicus.
Chloroplast.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids I; Fabiales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
NCBI_TaxID=34305;
(1)
SEQUENCE FROM N.A.
STRAIN=Accession MG-20;
MEDLINE=21082929; PubMed=11214967;
Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;
"Complete structure of the chloroplast genome of a legume, Lotus
japonicus";
DNA Res. 7:323-330 (2000).
-!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
of DNA into RNA using the four ribonucleoside triphosphates as
substrates.
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
{RNA} (N).
-!- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four
subunits: alpha, beta, beta', and beta".
-!- SUBCELLULAR LOCATION: Chloroplast.
-!- SIMILARITY: Belongs to the RNA polymerase beta chain family.
This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AF002983; BAB3194.1; -.
HSP; O9KMU7; 1HQW.
InterPro: IPR007121; RNA pol B.
InterPro: IPR007644; RNA pol Rpb2 1.
InterPro: IPR007642; RNA pol Rpb2 2.
InterPro: IPR007645; RNA pol Rpb2 3.
InterPro: IPR007120; RNA pol Rpb2 6.
InterPro: IPR007641; RNA pol Rpb2 7.
Pfam: PF04563; RNA pol Rpb2 1; 1.
Pfam: PF04561; RNA pol Rpb2 2; 1.
Pfam: PF04565; RNA pol Rpb2 3; 1.
Pfam: PF00562; RNA pol Rpb2 6; 1.
Pfam: PF04560; RNA pol Rpb2 7; 1.
PROSITE; PS01166; RNA POL BETA; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription; Chloroplast.
SQ SEQUENCE 1070 AA; 120892 MW; 7126BD99B0B5C063 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 1070;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 242 ESLCRELQ 249
Db 241 ESLCRELQ 248
RESULT 13
CARB_THETN STANDARD; PRT; 1072 AA.
ID CARB_THETN
AC Q8RBK0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
GN CARB OR TT0816.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;

Thermoanaerobacteriaceae; Thermoanaerobacter.
NCBI_TaxID=119072;
(1)
SEQUENCE FROM N.A.
STRAIN=M54 / JCM 11007;
MEDLINE=2192816; PubMed=11997336;
Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Bao Q., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
"A complete sequence of T. tengcongensis genome";
Genome Res. 12:689-700 (2002).
-!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate. (By similarity).
-!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
-!- PATHWAY: Arginine biosynthesis.
-!- SUBUNIT: Pyrimidine biosynthesis; first step.
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
-!- SIMILARITY: Belongs to the carb family.
This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AE013048; AAM24073.1; -.
HAMAP; MP 01210; -; 1.
InterPro: IPR006275; CarA_L_glu.
InterPro: IPR005483; CPase_L_D2.
InterPro: IPR005479; CPase_L_D3.
InterPro: IPR005480; CPase_L_D3.
InterPro: IPR005481; CPase_L_N.
InterPro: IPR004362; MGS like.
Pfam: PF00289; CPasease_L_Chain; 2.
Pfam: PF02786; CPasease_L_D2; 2.
Pfam: PF02787; CPasease_L_D3; 1.
Pfam: PF02142; MGS; 1.
PRINTS; PR00098; CPASE.
TIGRFAMs; TIGR01369; CPasease1_lrg; 1.
PROSITE; PS00866; CPASE 1; 1.
PROSITE; PS00867; CPASE 2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
FT DOMAIN 402 546
FT DOMAIN 547 929
FT DOMAIN 930 1072
FT REPEAT 1 546
FT REPEAT 547 1072
FT NP_BIND 153 210
FT NP_BIND 302 352
FT METAL 284 284
FT METAL 298 298
FT METAL 300 300
FT METAL 820 820
FT METAL 832 832
SQ SEQUENCE 1072 AA; 119000 MW; D5F08ADC8BE75D37 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 1072;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 254 SLKEEGVQ 261
Db 36 SLKEEGVQ 43
RESULT 14

```

TFRAA_AGR75          STANDARD;      PRT;   1100 AA.
TRAA AGRTS
Q44349;
01-NOV-1997 (Rel. 35, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Conjugal transfer protein traA.
TFAA OR A106127 OR AGR_FTI_237.
Agrobacterium tumefaciens (strain C58 / ATCC 33970).
Plasmid pRiC58.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales.
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
NCBI_TaxID=176299;
[1]
SEQUENCE FROM N.A.
MEDLINE=96312367; PubMed=8763953;
Farrand S.K., Hwang I., Cook D.M.;
"The tra region of the nopaline-type Ti plasmid is a chimera with elements related to the transfer systems of RSF1010, RP4, and F."; J. Bacteriol. 178:4233-4247(1996).
[2]
SEQUENCE FROM N.A.
MEDLINE=21608550; PubMed=11743193;
Wood D.W., Setubal J.C., Kaul R.J., Monks D.E., Kitajima J.P., Okura V.K., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L., Chen Y., Paulsen I.T., Bisen J.A., Karst P.D., Bovee D. Sr., Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C., Kutaycin T., Levy R., Li M.-J., McClelland E., Palmieri A., Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M., Chumley F., Tingey S.V., Tomb J.-P., Gordon M.P., Olson M.V., Nestor E.W.;
"The genome of the natural genetic engineer Agrobacterium tumefaciens C58."; Science 294:2317-2323(2001).
[3]
SEQUENCE FROM N.A.
MEDLINE=21608551; PubMed=11743194;
Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Roumel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F., Molam C., Allinger M., Dougherty D., Scott C., Lappas C., Markez B., Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G., Ciolo C., Slater S.;
"Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58."; Science 294:2323-2328(2001).
-- SIMILARITY: Belongs to the mobA/mobL family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
-----
ENBL; AF010180; AAC17212.1; -.
ENEL; AE009431; ALA46363.1; -.
ENEL; AE007939; AAK91091.1; ALT_INIT.
PIR; AE3243; AK3243.
PIR; T03419; T03419.
InterPro; IPR005053; MobA_MoBL.
PFam; PF03389; MobA_MoBL; 1.
Conjugation; ATP-binding; Plasmid; Complete proteome.
NP_BIND 404 411
PEPTIDE 372 372
CONFLICT 372 372
SEQUENCE 1100 AA; 12347 MW; B85D31DBS26B7344 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 1100;
Best Local Similarity 100.0%; Pref.No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;

```

FT DOMAIN 742 763 IQ 1.
FT DOMAIN 764 787 IQ 2.
FT DOMAIN 788 817 IQ 3.
FT DOMAIN 1206 1304 PH 1.
FT DOMAIN 1386 1491 PH 2.
FT DOMAIN 1694 2038 FERM.
FT DOMAIN 800 941 COILED COIL (POTENTIAL).
SQ SEQUENCE 2052 AA; 235837 MW; 43DF13424B4B2D28 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 2052;
Best Local Similarity 100.0%; Pred.No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 264 REEEKRX 271
|||
Db 824 REEEKRX 831

Search completed: June 8, 2004, 16:38:39
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:26:25 ; Search time 60 Seconds
(without alignments)
2571.182 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 2785

Sequence: 1 MNQDKKNGAKQSNPKSP.....APSTEASGQTGPQEPSARA 546

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2785	100.0	546	4	AAB82808	Human low
2	2785	100.0	539	6	ABR41210	Human DIT
3	2785	100.0	639	6	ABR41186	Human DIT
4	2693	96.7	530	2	AAW49042	Human low
5	2693	96.7	530	4	AAB82804	Human low
6	2552.5	91.7	557	2	AAW49039	Rabbit lo
7	2552.5	91.7	557	4	AAB82801	Rabbit lo
8	2430.5	87.3	1749	4	ABG00839	Novel hum
9	2416	86.8	510	4	ABB11764	Human LDL
10	2416	86.8	510	4	AAW79741	Human pro
11	1801	64.7	356	6	ABO07223	Human p53
12	1754.5	63.0	358	4	AAW78757	Human pro
13	1284.5	46.1	528	4	AAB93594	Human pro
14	1275	45.8	684	6	ABR43284	Human neu
15	1265.5	45.4	395	6	ABO07222	Human p53
16	1244	44.7	676	2	AAI17863	Neurite e
17	1039.5	37.3	386	2	AAW37883	BRCAL mod
18	1039.5	37.3	386	2	AAI30151	Amino aci
19	815	29.3	204	4	AAB63259	Human bre
20	810	29.1	475	4	AAO08301	Human pol
21	810	29.1	475	7	ADC33216	Human nov
22	659	23.7	416	7	ADC31613	Human nov
23	642.5	23.1	410	4	ABG02486	Novel hum
24	597.5	21.5	515	4	ABB61894	Drosophil
25	550	19.7	841	4	ABG02487	Novel hum

ALIGNMENTS

RESULT 1

AAB82808

ID AAB82808 standard; protein; 546 AA.

XX

AC AAB82808;

XX

DT 12-NOV-2001 (first entry)

XX

DE Human low density lipoprotein binding protein 3 (LBP-3).

DB

XX Low density lipoprotein binding protein 3; LBP-3; LDL; human;

KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.

XX

OS Homo sapiens.

XX

PN WO200164874-A2.

XX

PD 07-SEP-2001.

XX

PF 28-FEB-2001; 2001WO-US006356.

XX

PR 02-MAR-2000; 2000US-00517849.

XX

PR 14-JUL-2000; 2000US-00616289.

XX

PA (BOST-) BOSTON HEART FOUND INC.

XX

PI Lees AM, Lees RS, Law SW, Arjona AA;

XX

DR WPI; 2001-565505/63.

XX

DR N-PSDB; AAH26501.

XX

PT New isolated low density lipoprotein binding polypeptide for treating, diagnosing and/or identifying therapeutic agents for atherosclerosis.

XX

PF Claim 13(1); Fig 8A; 143pp; English.

XX

CC The present sequence is that of novel human low density lipoprotein binding protein 3 (LBP-3). The amino acid sequence was deduced from the coding region of isolated genomic DNA (see AAH26501). It differs from the sequence predicted from an LBP-3 cDNA clone (see AAB82804) by the presence of an additional 16 amino acids at the N-terminus (the cDNA clone is 5' truncated) and by having asparagine at amino acid position 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP polypeptides of the invention that are capable of binding to native and methylated low density lipoproteins. Also claimed are biologically active fragments and analogues of these LBPs, polynucleotides encoding LBPs, as well as expression vectors, cells and methods of producing the LBPs. A polypeptide having the amino acid residues 96-110 of the present sequence

26	322	11.6	102	4	ABG02488	Novel hum
27	291	10.4	274	4	ABB58657	Drosophil
28	278.5	10.0	1455	7	ADSB0783	Microsate
29	275.5	9.9	954	4	AAU14615	Novel bon
30	263	9.4	1752	2	AAI07031	Breast ca
31	263	9.4	2871	6	ABU07402	Protein d
32	263	9.4	2871	7	ADC35075	Human bre
33	263	9.4	2918	4	ABG27218	Novel hum
34	260	9.3	909	3	AAB53448	Human col
35	258	9.3	1206	7	ADD46507	Rat Prote
36	258	9.3	1206	7	ADD46507	Rat Prote
37	257.5	9.2	860	7	ADC38517	Human AML
38	255.5	9.2	931	4	AAW79504	Human pro
39	255.5	9.2	931	4	AAW79504	Human pro
40	255	9.2	1240	4	AGG67538	Amino aci
41	254.5	9.1	1880	4	AAB96332	Putative
42	253.5	9.1	1484	2	AAH89721	Canine ri
43	253.5	9.1	1690	4	ABB61144	Drosophil
44	253.5	9.1	1690	4	ABB61173	Drosophil
45	252.5	9.1	1881	5	ABP73809	Candida a

Db 94 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEQA 153
Qy 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAEKSRITYARN 120
Db 154 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAEKSRITYARN 213
Qy 121 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKXKAKGLGKITTLLMQTLN 180
Db 214 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKXKAKGLGKITTLLMQTLN 273
Qy 181 TLSTPEKLAALCKKYAELLBEHRNSQKMKLQKQSQLVQEKDHLRGHSHKAVLARSK 240
Db 274 TLSTPEKLAALCKKYAELLBEHRNSQKMKLQKQSQLVQEKDHLRGHSHKAVLARSK 333
Qy 241 LESLCELQHNRSKLEEGVQAREREEKKEVTSHPQVTLNDIQLQMEQHNRNSKLQR 300
Db 334 LESLCELQHNRSKLEEGVQAREREEKKEVTSHPQVTLNDIQLQMEQHNRNSKLQR 393
Qy 301 ENWELAEALRLKLTIEQVELREHIDKVFVKHKOQQQLVDKLAQQAQSMLEAEERHQREK 360
Db 394 ENWELAEALRLKLTIEQVELREHIDKVFVKHKOQQQLVDKLAQQAQSMLEAEERHQREK 453
Qy 361 FLKKEAVESQRCMELMKQETHLKOALALYTKFEFEFONTLSSSEVFTTPKQEMERMTK 420
Db 454 FLKKEAVESQRCMELMKQETHLKOALALYTKFEFEFONTLSSSEVFTTPKQEMERMTK 513
Qy 421 KIKLEKETTMVRSRNESSKALLENMAEKTVRDKELEGLOVKIQRLKLCRALQTERND 480
Db 514 KIKLEKETTMVRSRNESSKALLENMAEKTVRDKELEGLOVKIQRLKLCRALQTERND 573
Qy 481 LNKRVODLSAGGQSGSLTDSGPERPBGQAAPSSPRVTEAPCYGAPSTEASGQTGPQE 540
Db 574 LNKRVODLSAGGQSGSLTDSGPERPBGQAAPSSPRVTEAPCYGAPSTEASGQTGPQE 633
Qy 541 PTSARA 546
Db 634 PTSARA 639

RESULT 3
ABR41186
ID ABR41186 standard; protein; 639 AA.
XX AC ABR41186;
XX DT 02-JUN-2003 (first entry)
DE Human DITHP extracellular signalling protein.
XX Human: dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW extracellular signalling.
XX OS Homo sapiens.
XX WO200297031-A2.
XX PD 05-DEC-2002.
XX PF 27-MAR-2002; 2002WO-US010056.
XX PR 28-MAR-2001; 2001US-0279619P.
XX PR 29-MAR-2001; 2001US-0280067P.
XX PR 16-MAY-2001; 2001US-0291280P.
XX PR 17-MAY-2001; 2001US-0291829P.
XX PR 17-MAY-2001; 2001US-0291849P.
XX PR 19-JUN-2001; 2001US-0299428P.

PR 20-JUN-2001; 2001US-0299776P.
XX 20-JUN-2001; 2001US-0300001P.
PA (INCY-) INCYTE GENOMICS INC.
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marvaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
DR N-PSDB; ACC46130.
XX
PT Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS Claim 27; SEQ ID NO 721; 591pp; English.
XX
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; gastrointestinal disorders; transport
CC disorders; neurological disorders; hormonal disorders; metabolic
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has extracellular
CC signalling activity. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 639 AA;
Query Match 100.0%; Score 2785; DB 6; Length 639;
Best Local Similarity 100.0%; Pred. No. 6.8e-160; Indels 0; Gaps 0;
Matches 546; Conservative 0; Mismatches 0;
Qy 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEQA 60
Db 94 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEQA 153
Qy 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAEKSRITYARN 120
Db 154 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAEKSRITYARN 213
Qy 121 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKXKAKGLGKITTLLMQTLN 180
Db 214 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKXKAKGLGKITTLLMQTLN 273
Qy 181 TLSTPEKLAALCKKYAELLBEHRNSQKMKLQKQSQLVQEKDHLRGHSHKAVLARSK 240
Db 274 TLSTPEKLAALCKKYAELLBEHRNSQKMKLQKQSQLVQEKDHLRGHSHKAVLARSK 333
Qy 241 LESLCELQHNRSKLEEGVQAREREEKKEVTSHPQVTLNDIQLQMEQHNRNSKLQR 300

PR	14-JUL-2000; 2000US-00616289.	AAW49039 standard; protein; 557 AA.
XX	(BOST-) BOSTON HEART FOUND INC.	AAW49039;
PA	Lees AM, Lees RS, Law SW, Arjona AA;	09-NOV-1998 (first entry)
PI	WPI; 2001-565505/63.	Rabbit low density lipoprotein binding protein LBP-3.
DR	N-PSDB; AAW26496.	Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
XX	New isolated low density lipoprotein binding polypeptide for treating,	receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.
PT	diagnosing and/or identifying therapeutic agents for atherosclerosis.	
PS	Claim 13(h); Fig 8B; 143pp; English.	
XX	The present sequence is that of the N-terminal portion of novel human low	
CC	density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is	
CC	deduced from an isolated partial cDNA clone (see AAW26494). A full-length	
CC	sequence is given in AAW2808. Human LBP-3 is an example of claimed LBP	
CC	polypeptides of the invention that are capable of binding to native and	
CC	methylated low density lipoproteins. Also claimed are biologically active	
CC	fragments and analogues of LBPs, polynucleotides encoding LBPs, as well	
CC	as expression vectors, cells and methods of producing the LBPs. Methods	
CC	of determining if an animal is at risk for atherosclerosis, methods for	
CC	evaluating an agent for use in treating atherosclerosis, and methods for	
CC	treating a cell having an abnormality in structure or metabolism of LBP	
CC	are claimed. Pharmaceutical compositions comprising an LBP polypeptide or	
CC	nucleic acid, and vaccine compositions, are also claimed	
XX	Sequence 530 AA;	
SQ	Query Match 96.7%; Score 2693; DB 4; Length 530;	
	Best Local Similarity 99.8%; Pred. No. 2e-154; Indels 0; Gaps 0;	
	Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	17 KSSPGQPEAGPEGAQERPSQAAPVAEAGSGSSQAQPRKPEGAQTAQSGALRDVSEELS 76	
DB	1 KSSPGQPEAGPEGAQERPSQAAPVAEAGSGSSQAQPRKPEGAQTAQSGALRDVSEELS 60	
QY	77 RQLEDILSTYCVNNOGCGEDGAGPEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 136	
DB	61 RQLEDILSTYCVNNOGCGEDGAGPEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 120	
QY	137 KGPNTPEIRQSDVEGDRHRRPQEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKKY 196	
DB	121 KGPNTPEIRQSDVEGDRHRRPQEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKKY 180	
QY	197 AELLEHRNSQKQKLLQKQSLVQKQHLRQEHKAVLARSKLSLCLRELQHRNSLK 256	
DB	181 AELLEHRNSQKQKLLQKQSLVQKQHLRQEHKAVLARSKLSLCLRELQHRNSLK 240	
QY	257 EGVQVAREEERKKEVTSHFQVTLNDIQLQMHQHRNSKLAQENNELAERLKLIEQY 316	
DB	241 EGVQVAREEERKKEVTSHFQVTLNDIQLQMHQHRNSKLAQENNELAERLKLIEQY 300	
QY	317 ELREEHIDKVFKHDLQOQLVDAKLQQAQEMLKAEERHOREKDFLLKEAVESQRMCELM 376	
DB	301 ELREEHIDKVFKHDLQOQLVDAKLQQAQEMLKAEERHOREKDFLLKEAVESQRMCELM 360	
QY	377 KQETHLKOQALYTERPEFQNTLSKSVFTTFOEMKMTKKLCKLEKETTMYRSRW 436	
DB	361 KQETHLKOQALYTERPEFQNTLSKSVFTTFOEMKMTKKLCKLEKETTMYRSRW 420	
QY	437 ESSNKALLEMAEKTVDKELSGLOVKIQRLEKLCRALQTERNDLNKRVQDLGAGOGSL 496	
DB	421 ESSNKALLEMAEKTVDKELSGLOVKIQRLEKLCRALQTERNDLNKRVQDLGAGOGSL 480	
QY	497 TDSPGPRPFGQAQSSPRVTAPCYPCAPSTASGQGPQEPPTSARA 546	
DB	481 TDSPGPRPFGQAQSSPRVTAPCYPCAPSTASGQGPQEPPTSARA 530	
RESULT 6		
AAW49039		

ID	AAW49039 standard; protein; 557 AA.	
XX	AAW49039;	
AC	09-NOV-1998 (first entry)	
XX	Rabbit low density lipoprotein binding protein LBP-3.	
DE	Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;	
XX	receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.	
KW	Oryctolagus cuniculus.	
OS		
XX	Key Location/Qualifiers	
FT	Peptide 96..110	
FT	/note= "Claim 2"	
XX	WO9823282-A1.	
PN	04-JUN-1998.	
XX	26-NOV-1997; 97WO-US021857.	
PD	27-NOV-1996; 96US-0031930P.	
XX	03-JUN-1997; 97US-0048547P.	
PR	(BOST-) BOSTON HEART FOUND INC.	
BR		
XX	Lees AM, Lees RS, Law SW, Arjona AA;	
PA	WPI; 1998-322455/28.	
XX	N-PSDB; AAV32836.	
DR	Nucleic acid encoding low density lipoprotein binding proteins and	
XX	related vectors - transformed cells, proteins, and modulators of binding,	
PT	useful for treatment and diagnosis of atherosclerosis and for identifying	
PT	subjects at risk.	
XX	Claim 1; Fig 5; 47pp; English.	
PS	This polypeptide comprises novel rabbit low density lipoprotein (LDL)	
CC	binding protein LBP-3 that is capable of binding both native and methyl	
CC	LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA	
CC	(see AAV32836). cDNA clones (see AAV32834-39) and encoded rabbit and human	
CC	LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP	
CC	metabolism or structure is diagnostic of a risk for atherosclerosis. The	
CC	invention provides: methods for determining if an animal is at risk for	
CC	atherosclerosis (e.g. for prenatal screening); methods for treating	
CC	atherosclerosis (including gene therapy) using e.g. LBP polypeptides to	
CC	bind LDL and thereby prevent formation of atherosclerotic plaque; and	
CC	methods for treating a cell having an abnormality in LBP structure or	
CC	metabolism. Pharmaceutical and vaccine compositions are also provided, as	
CC	well as recombinant vectors and host cells used to produce recombinant	
CC	LBP	
XX	Sequence 557 AA;	
SQ	Query Match 91.7%; Score 2552.5; DB 2; Length 557;	
	Best Local Similarity 91.0%; Pred. No. 6.2e-146;	
	Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;	
QY	1 MKNQDKKNGAAQKSNPKSPGQPEAGPEGAQERPSQAAPVAEAGSGSSQAQPRKPEGAQ 60	
DB	1 MKNQDKKNGAAQKSNPKSPGQPEAGPEGAQERPSQAAPVAEAGSGSSQAQPRKPEGAQ 59	
QY	61 RTAQSGALRDVSEELSRLQLEDILSTYCVNNOGCGEDGAGPEPAEPAEPAEPAEPAE 120	
DB	60 RTAQSGALRDVSEELSRLQLEDILSTYCVNNOGCGEDGAGPEPAEPAEPAEPAEPAE 119	
QY	121 GSEPEP-TPVWNGEKPKSGDPNTEHIRQSDVEGDRHRRPQEKKAQGLGKEITLLMOTL 179	
DB	120 GSEPEP-TPVWNGEKPKSGDPNTEHIRQSDVEGDRHRRPQEKKAQGLGKEITLLMOTL 179	

180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSOVLQVQKHILRGEHSHKAVLARS 239
 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSOVLQVQKHILRGEHSHKAVLARS 239
 240 KLESICRELQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
 240 KLESICRELQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
 300 QENNELAERLKLLEQYELREHIDKVFKHDLQOQLVDKLLQQAQEMKLEAEERHOREK 359
 300 QENNELAERLKLLEQYELREHIDKVFKHDLQOQLVDKLLQQAQEMKLEAEERHOREK 359
 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKEFEFQNTLSKSSSEVFTTFKQEMKMT 419
 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKEFEFQNTLSKSSSEVFTTFKQEMKMT 419
 420 KKI KLEKETTYMYSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
 420 KKI KLEKETTYMYSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGPAP 528
 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGPAP 528
 529 STEASGQTGPQEPPTSARA 546
 540 STEASGQTGPQEPPTSARA 557

RESULT 7
 ID AAB82801 standard; protein; 557 AA.
 AC AAB82801;
 XX 12-NOV-2001 (first entry)
 XX Rabbit low density lipoprotein binding protein 3 (LBP-3).
 DE Low density lipoprotein binding protein 3; LBP-3; LDL; rabbit;
 KW atherosclerosis; antiatherosclerotic; therapy; diagnosis; vaccine.
 XX Oryctolagus cuniculus.
 XX WO200164874-A2.
 XX 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-US006356.
 XX 02-MAR-2000; 2000US-00517849.
 XX 14-JUL-2000; 2000US-00616289.
 XX (BOST-) BOSTON HEART FOUND INC.
 XX Lees AM, Lees RS, Law SW, Arjona AA;
 XX WPI; 2001-565505/53.
 XX N-PSDB; AAB26491.

New isolated low density lipoprotein binding polypeptide for treating, diagnosing and/or identifying therapeutic agents for atherosclerosis.
 Claim 13(e); Fig 5; 143pp; English.
 The present sequence is that of novel rabbit low density lipoprotein binding protein 1 (LBP-3). The amino acid sequence is deduced from an isolated cDNA clone (see AAB26491). Rabbit LBP-3 is an example of claimed polypeptides of the invention, termed LBPs, that are capable of binding to native and methylated low density lipoproteins. Also claimed are biologically active fragments and analogues of LBPs, polynucleotides encoding LBPs, as well as expression vectors, cells and methods of producing the LBPs. A polypeptide having the amino acid residues 96-110

CC of the present sequence is claimed (see AAB82819). Methods of determining if an animal is at risk for atherosclerosis, methods for evaluating an agent for use in treating atherosclerosis, and methods for treating a cell having an abnormality in structure or metabolism of LBP are claimed.
 CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic acid, and vaccine compositions, are also claimed
 XX
 SQ Sequence 557 AA;

Query Match 91.7%; Score 2552.5; DB 4; Length 557;
 Best local Similarity 91.0%; Pred. No. 6.2e-146;
 Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;
 QY 1 MNQDKKNGAAKQNPKNKSPGQPEAGPEGAQEPSPQAAPAVEAGPGSSQAPRKPEGAQA 60
 DB 1 MNQDKKNGAAKQNPKNKSPGQPEAGPEGAQEPSPQAAPAVEAGPGSSQAPRKPEGAQA 59
 QY 61 RTQSGALRDVSELRQLEDILSTCYVDNQQGPGEDGAQEPAPPAEAKKSTTVARN 120
 DB 60 KTAQPGALCDVSELRQLEDILSTCYVDNQQGPGEDGVQGEPPPEDAKSRAYARN 119
 QY 121 GEPEP-TPVUNGKEPSKGPNTPEIRQSDVEGDRHRRPQEKKAKGLCKEITLLMQTL 179
 DB 120 GEPEP-TPVUNGKEPSKGPNTPEIRQSDVEGDRHRRPQEKKAKGLCKEITLLMQTL 179
 QY 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSOVLQVQKHILRGEHSHKAVLARS 239
 DB 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSOVLQVQKHILRGEHSHKAVLARS 239
 QY 240 KLESICRELQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
 DB 240 KLESICRELQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
 QY 300 QENNELAERLKLLEQYELREHIDKVFKHDLQOQLVDKLLQQAQEMKLEAEERHOREK 359
 DB 300 QENNELAERLKLLEQYELREHIDKVFKHDLQOQLVDKLLQQAQEMKLEAEERHOREK 359
 QY 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKEFEFQNTLSKSSSEVFTTFKQEMKMT 419
 DB 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKEFEFQNTLSKSSSEVFTTFKQEMKMT 419
 QY 420 KKI KLEKETTYMYSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
 DB 420 KKI KLEKETTYMYSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
 QY 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGPAP 528
 DB 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGPAP 528
 QY 529 STEASGQTGPQEPPTSARA 546
 DB 540 STEASGQTGPQEPPTSARA 557

RESULT 8
 AAB82819
 ID AEG00839 standard; protein; 1749 AA.
 AC AEG00839;
 XX 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #830.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX

PF 30-MAR-2001; 2001WO-US028631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSB-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS65026.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 31198; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1749 AA;

Query Match 87.3%; Score 2430.5; DB 4; Length 1749;
Best Local Similarity 84.9%; Pred. No. 5.3e-138;
Matches 500; Conservative 4; Mismatches 30; Indels 55; Gaps 6;

Qy 1 MNQDKKNGAAKQSNPKSSPCQPEAGPEGAGQERPSQAAPAVEAGPGSSQAPRKPEGAQA 60
Db 152 MNQDKKNGAAKQSNPKSSPCQPEAGPEGAGQERPSQAAPAVEAGPGSSQAPRKPEGAQA 211

Qy 61 RTAQSGLRDVSELSRQLEDILSTYCVNNQGGGEGDGAQGEPAEPDEAKERTYVARN 120
Db 212 RTAQSGLRDVSELSRQLEDILSTYCVNNQGGGEGDGAQGEPAEPDEAKERTYVARN 271

Qy 121 GEPEPTPWNGEKPSSKGDPTNTEIRSDVSGDRHRRPQEKKAKGL-----GK 170
Db 272 GEPEPTPWNGEKPSSKGDPTNTEIRSDVSGDRHRRPQEKKAKGLGEPRAALCEAGK 331

Qy 171 EITLLMOTLNTLSTPEBKLAALCKKYAELIEHNSOKMKLLQKQSQQLVQEKDHURGE 230
Db 332 EITLLMOTLNTLSTPEBKLAALCKKYAELIEHNSOKMKLLQKQSQQLVQEKDHURGE 391

Qy 231 HSKAVLARSKLESICRELRNRSIK-----PEGVQARERREKR 270
Db 392 HSKAVLARSKLESICRELRNRSIKLGDPCVAVDDGSRADPFAQEGEVQARERREKR 451

Qy 271 KEVTSHEQVTLNDILQMQEQHNSRNSKLRQENMELAEKLIQVLEEEHIDKVEFKH 330
Db 452 KEVTSHEQVTLNDILQMQEQHNSRNSKLRQENMELAEKLIQVLEEEHIDKVEFKH 511

Qy 331 DLQQLVDKLAQQAQEMLKAEERHQRKDFLLKEAVESQRMCELMKQQTTHLKKQALY 390

Db 512 DLQQLVDKLAQQAQEMLKAEERHQRKDFLLKEAVESQRMCELMKQQTTHLKKQALY 571
Qy 391 TEKTFEPQNTLSKSSEVFTTFQEKEMTKKIKKLEKETTMYRSRWSSENKALLEAEK 450
Db 572 TEKTFEPQNTLSKSSEVFTTFQEKEMTKKIKKLEKETTMYRSRWSSENKALLEAEK 631
Qy 451 TVRDXELEGLQVKIQRIELEKLCRALQTERNDLNKRVDL-----SAGGQGSILT 497
Db 632 TVRDXELEGLQVKIQRIELEKLCRALQTERNDLNKRVDL-----SAGGQGSILT 691
Qy 498 DSGPRRRPREGQAQAPSS-----PRVTEAPCYPGAPSTASQGTGPQEP 541
Db 692 BEALSRCPOEQCFMLPFTSAKRNPR---KPLLPF-----DGEGLRPSQRP 733

RESULT 9
ABBI1764
ID ABBI1764 standard; peptide; 510 AA.
XX
AC ABBI1764;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human Lf6 binding protein homologue, SEQ ID NO:2134.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;
KW antifungal; vulnery; antitumor.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSB-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457740/49.
DR N-PSDB; ABA09008.
XX
XX Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX
PS Claim 20; Page 240-241; 1963pp; English.
XX
CC Sequences ABBI10981-ABBI12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides,
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may

CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel, in drug
 CC polypeptide of the invention

XX SQ Sequence 510 AA;

Query Match 86.8%; Score 2416; DB 4; Length 510;
 Best Local Similarity 100.0%; Pred. No. 9.7e-138;
 Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAKQSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 60
 DB 12 MNQDKKNGAAKQSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 71
 QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 120
 DB 72 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 131
 QY 121 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKILTMQTLN 180
 DB 132 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKILTMQTLN 191
 QY 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLQKQSQLVQEKDLRGESKAVLARSK 240
 DB 192 TLSTPEEKLAALCKKYAEELLEHRNSQKMLQKQSQLVQEKDLRGESKAVLARSK 251
 QY 241 LESLCRELQNRNSLKEGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
 DB 252 LESLCRELQNRNSLKEGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQ 311
 QY 301 ENMELAEERLKKLIQYELAREHHDKVPFKHQLQQLQVDAKQQQAQEMKAEERHQRKD 360
 DB 312 ENMELAEERLKKLIQYELAREHHDKVPFKHQLQQLQVDAKQQQAQEMKAEERHQRKD 371
 QY 361 FLKKAVERQRMCELMKQETHLKKQALATYRKFEFFQNTLSKSEVFTTFQEMKMTK 420
 DB 372 FLKKAVERQRMCELMKQETHLKKQALATYRKFEFFQNTLSKSEVFTTFQEMKMTK 431
 QY 421 KIKKLEKETTMYRSRWSSNKALLEWABEKTVRDKLEGLQVKIQRLEKLCRALQT 476
 DB 432 KIKKLEKETTMYRSRWSSNKALLEWABEKTVRDKLEGLQVKIQRLEKLCRALQT 487

RESULT 10

AAM79741

ID AAM79741 standard; protein; 510 AA.

XX AC

XX AAM79741;

XX

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3387.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;
 XX nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

XX W0200157190-A2.

PD 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00568875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00628325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSB-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX N-PSDB; AAKS2874.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

XX in diagnosis and gene therapy.

XX Claim 20; Page 319; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 XX cytokine, cell proliferation or cell differentiation or which may induce
 XX production of other cytokines in other cell populations. The
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
 XX peptide therapy. The polypeptides have various cytokine-like activities,
 XX e.g. stem cell growth factor activity, haematopoiesis regulating
 XX activity, tissue growth factor activity, immunomodulatory activity and
 XX activin/inhibin activity and may be useful in the diagnosis and/or
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
 XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 XX (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 XX sequence listing were missing at the time of publication

XX Sequence 510 AA;

Query Match 86.8%; Score 2416; DB 4; Length 510;
 Best Local Similarity 100.0%; Pred. No. 9.7e-138;
 Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAKQSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 60
 DB 12 MNQDKKNGAAKQSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 71
 QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 120
 DB 72 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 131
 QY 121 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKILTMQTLN 180
 DB 132 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKILTMQTLN 191
 QY 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLQKQSQLVQEKDLRGESKAVLARSK 240

DB 192 TLSTPEKLAALCKYAELEHNSQKMLQKQSQVLQKDHGHSKAVLARSK 251
QY 241 LESLCRELQNRSLKEEGVQVAREBEKKEVTSHPQVTLNDIQLQMEQHNRNSKLARQ 300
DB 252 LESLCRELQNRSLKEEGVQVAREBEKKEVTSHPQVTLNDIQLQMEQHNRNSKLARQ 311
QY 301 ENMELAEKLLIYOYELREEHIDKVPKHDLQQLVDKLOQAQEWLKEAEERHOREKD 360
DB 312 ENMELAEKLLIYOYELREEHIDKVPKHDLQQLVDKLOQAQEWLKEAEERHOREKD 371
QY 361 FLKEAVESQRMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTX 420
DB 372 FLKEAVESQRMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTX 431
QY 421 KIKKLEKETTMYRSRWNSSNKALLEWAEKTVRDKEGLQVKIQRLKLCRALQ 476
DB 432 KIKKLEKETTMYRSRWNSSNKALLEWAEKTVRDKEGLQVKIQRLKLCRALQ 487

RESULT 11
ABO07223
ID ABO07223 standard; protein; 356 AA.
XX AC ABO07223;
XX DT 13-AUG-2003 (first entry)
XX DE Human p53 modifying protein, SEQ ID 183.
XX KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
XX KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
XX KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
XX KW apoptotic disorder; cell proliferation disorder.
XX OS Homo sapiens.
XX PN W0200299122-A1.
XX PD 12-DEC-2002.
XX PF 03-JUN-2002; 2002WO-US017382.
XX PR 05-JUN-2001; 2001US-0296076P.
XX PR 10-OCT-2001; 2001US-0328605P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX WPI; 2003-156859/15.
XX DR N-PSDB; ACD13448.
XX PT Identifying modulators of the p53 pathway for use in treating apoptotic
XX PT or cell proliferation disorders, comprises screening for agents that
XX PT modulate activity of a human ortholog of genes that modify the p53
XX PT pathway in Drosophila.

XX PS Example 2; Page 533-534; 678pp; English.
XX CC The invention relates to identifying (M1) a candidate p53 pathway
XX CC modulating agent, by contacting an assay system comprising a purified HM
XX CC polypeptide (human orthologue of genes that modify the p53 pathway in
XX CC Drosophila) or nucleic acid with a test agent under conditions, where but
XX CC for the presence of the test agent, the system provides a reference
XX CC activity, and detecting a test agent-biased activity of the assay system.
XX CC Also included are modulating (M2) a p53 pathway of a cell (comprising
XX CC contacting a cell defective in p53 function with a candidate modulator
XX CC that specifically binds to a HM polypeptide comprising an HM amino acid
XX CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
XX CC in a mammalian cell (comprising contacting the cell with an agent that
XX CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)

CC a disease in a patient (comprising: (a) obtaining a biological sample
CC from the patient; (b) contacting the sample with a probe for HM
CC expression; (c) comparing the results with a control; and (d) determining
CC whether the comparison indicates a likelihood disease). (M1) is useful
CC for identifying modulators of the p53 pathway. A probe for HM expression
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC in a patient, where the cancer has greater than 25 % expression level.
CC Modulators identified by (M1) are useful in a variety of diagnostic and
CC therapeutic applications, where disease or disorder prognosis is related
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC the p53 function of the cell, so that the cell undergoes normal
CC proliferation or progression through the cell cycle. (M2) and (M3) are
CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein
XX
XX Sequence 356 AA;

Query Match 64.7%; Score 1801; DB 6; Length 356;
Best Local Similarity 100.0%; Pred. No. 8.3e-101;
Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 ALCKKYAELEHNSQKMLQKQSQVLQKDHGHSKAVLARSKLSLCELQ 250
DB 1 ALCKKYAELEHNSQKMLQKQSQVLQKDHGHSKAVLARSKLSLCELQ 60
QY 251 HNSLKEGVQVAREBEKKEVTSHPQVTLNDIQLQMEQHNRNSKLQENNELAERLK 310
DB 61 HNSLKEGVQVAREBEKKEVTSHPQVTLNDIQLQMEQHNRNSKLQENNELAERLK 120
QY 311 KLIEQYELAREHIDKVPKHDLQQLVDKLOQAQEWLKEAEERHOREKDPLLKEAVESQ 370
DB 121 KLIEQYELAREHIDKVPKHDLQQLVDKLOQAQEWLKEAEERHOREKDPLLKEAVESQ 180
QY 371 RMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTXKIKKLEKETT 430
DB 181 RMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTXKIKKLEKETT 240
QY 431 MYRSRWNSSNKALLEWAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVODLSA 490
DB 241 MYRSRWNSSNKALLEWAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVODLSA 300
QY 491 GGQGSILTDSGPERPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPOEPTSARA 546
DB 301 GGQGSILTDSGPERPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPOEPTSARA 356

RESULT 12
AAM78757
ID AAM78757 standard; protein; 358 AA.
XX AC AAM78757;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein SEQ ID NO 1419.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation.
XX OS Homo sapiens.
XX PN W0200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US004098.
XX PR 03-FEB-2000; 2000US-00496914.

27-APR-2000; 2000US-00560875.
20-JUN-2000; 2000US-00598075.
19-JUL-2000; 2000US-00620325.
01-SEP-2000; 2000US-00654936.
15-SEP-2000; 2000US-00663561.
20-OCT-2000; 2000US-00693325.
30-NOV-2000; 2000US-00728422.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
Ma Y, Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;
WPI; 2001-476283/51.
N-PSDB; AAK51890.
Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
Claim 20; Page 3681-3682; 6221pp; English.
The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM70323-AAM80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation. Note: Records for SEQ ID NO.2110 (AAK52581), 2111
(AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
sequence listing were missing at the time of publication
Sequence 358 AA;
Query Match 63.0%; Score 1754.5; DB 4; Length 358;
Best Local Similarity 98.3%; Pred. No. 5.4e-98;
Matches 349; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 193 CKKY-ABELEHNSQKQMLLQKQQLVQKQHLRGEHSAVLARSKLSLCLRLQRH 251
DB 4 CFPYPCQLEHNSQKQMLLQKQQLVQKQHLRGEHSAVLARSKLSLCLRLQRH 63
QY 252 NRSLEKGVQAREREKKEVTSFQVTLNDIQLOMEQHNERSKLRQENWELAEELKK 311
DB 64 NRSLEKGVQAREREKKEVTSFQVTLNDIQLOMEQHNERSKLRQENWELAEELKK 123
QY 312 LIEQVELREEHIDKVFKKHQLQQQLVDKQLQAQEMKLEAEERHQREKDFLLKEAVESQR 371
DB 124 LIEQVELREEHIDKVFKKHQLQQQLVDKQLQAQEMKLEAEERHQREKDFLLKEAVESQR 183
QY 372 MCELMKQOETHLKOQLALYTEKPEFQNTLSKSESVFTTFQEMKMTKKIKLEKFTTM 431
DB 184 MCELMKQOETHLKOQLALYTEKPEFQNTLSKSESVFTTFQEMKMTKKIKLEKFTTM 243
QY 432 YRSRVESNKALLEMAEKTVRDKEGLQVKIQLEKLCRALQTERNDLNKRVQDLSAG 491
DB 244 YRSRVESNKALLEMAEKTVRDKEGLQVKIQLEKLCRALQTERNDLNKRVQDLSAG 303
QY 492 CGGSITDSGPERPRPGGAQAPSPRVTEAPCYGAPSTEASGQTGPQEPPTSARA 546
DB 304 CGGSITDSGPERPRPGGAQAPSPRVTEAPCYGAPSTEASGQTGPQEPPTSARA 358
RESULT 13
AAB93594
ID AAB93594 standard; protein; 528 AA.
XX
AC AAB93594;
XX

26-JUN-2001 (first entry)
Human protein sequence SEQ ID NO:13026.
Human; primer; detection; diagnosis; antisense therapy; gene therapy.
Homo sapiens.
EP1074617-A2.
07-FEB-2001.
28-JUL-2000; 2000EP-00116126.
29-JUL-1999; 99JP-00248036.
27-AUG-1999; 99JP-00300253.
11-JAN-2000; 2000JP-00118776.
02-MAY-2000; 2000JP-00183767.
09-JUN-2000; 2000JP-00241899.
(HELI-) HELIX RES INST.
Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.
Primer sets for synthesising polynucleotides, particularly the 5602 full-
length cDNAs defined in the specification, and for the detection and/or
diagnosis of the abnormality of the proteins encoded by the full-length
cDNAs.
Claim 8; SEQ ID NO 13026; 2537pp + Sequence Listing; English.
The present invention describes primer sets for synthesising 5602 full-
length cDNAs defined in the specification. Where a primer set comprises:
(a) an oligo-dT primer and an oligonucleotide complementary to the
complementary strand of a polynucleotide which comprises one of the 5602
nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in the
specification. The primer sets can be used in antisense therapy and in
gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAK03166 to AAK13628 and
AAH13633 to AAK18742 represent human cDNA sequences; AAB92446 to AAB95893
represent human amino acid sequences; and AAK13629 to AAK13632 represent
oligonucleotides, all of which are used in the exemplification of the
present invention
Sequence 528 AA;
Query Match 46.1%; Score 1284.5; DB 4; Length 528;
Best Local Similarity 53.4%; Pred. No. 1.9e-59;
Matches 285; Conservative 75; Mismatches 131; Indels 43; Gaps 10;
QY 28 EGACQEPSQAPVAVRAGCGSSQAPK--PEGAQRTAGSALRDYSELSRQLEDTLST 85
DB 7 EAARGGGGAEEATEAGRGRRRSPQKFEIGTMEBAGICGLGVKADMLCNSQSDILQH 66
QY 86 YCVDDNQGPGEDGACQCEPAEPDAKSRITYVARNGEPEPTPVVNGEKSPKGPNTTEEI 145
DB 67 Q--GSNCGTGNKHSLEEDGSDFTYENMLVS-----PAYCTQSRREIPGG---EA 114
QY 146 ROSDEVGDRHRPRPQEKKAKGLGKGIITLLMQTILNTLSPEEKALALCKKYAELEHRN 205
XX

QY 541 PTSA 544
| : |
Db 574 PSKA 577

RESULT 15
ABC07222
ID ABO07222 standard; protein; 395 AA.

XX ABC07222;
XX 13-AUG-2003 (first entry)

XX Human p53 modifying protein, SEQ ID 182.

XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
XX antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
XX lung cancer; ovarian cancer; angiogenesis; cell cycle;
XX apoptotic disorder; cell proliferation disorder.

XX Homo sapiens.

XX WO200299122-A1.

XX 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017382.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156859/15.

XX N-PSDB; ACD13396.

XX Identifying modulators of the p53 pathway for use in treating apoptotic
XX or cell proliferation disorders, comprises screening for agents that
XX modulate activity of a human ortholog of genes that modify the p53
XX pathway in Drosophila.

XX Example 2; Page 532-533; 678pp; English.

XX The invention relates to identifying (M1) a candidate p53 pathway
XX modulating agent, by contacting an assay system comprising a purified HM
XX polypeptide (human orthologue of genes that modify the p53 pathway in
XX Drosophila) or nucleic acid with a test agent under conditions, where but
XX for the presence of the test agent, the system provides a reference
XX activity, and detecting a test agent-biased activity of the assay system.
XX Also included are modulating (M2) a p53 pathway of a cell (comprising
XX contacting a cell defective in p53 function with a candidate modulator
XX that specifically binds to a HM polypeptide comprising an HM amino acid
XX sequence, where p53 function is restored), modulating (M3) a p53 pathway
XX in a mammalian cell (comprising contacting the cell with an agent that
XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
XX a disease in a patient (comprising: (a) obtaining a biological sample
XX from the patient; (b) contacting the sample with a probe for HM
XX expression; (c) comparing the results with a control; and (d) determining
XX whether the comparison indicates a likelihood disease). (M1) is useful
XX for identifying modulators of the p53 pathway. A probe for HM expression
XX is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
XX in a patient, where the cancer has greater than 25 % expression level.
XX Modulators identified by (M1) are useful in a variety of diagnostic and
XX therapeutic applications, where disease or disorder prognosis is related
XX to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
XX proliferation disorders (e.g. cancer). Another two new methods (M2 and
XX M3) are useful for modulating the p53 pathway of a cell, thus restoring
XX the p53 function of the cell, so that the cell undergoes normal
XX proliferation or progression through the cell cycle. (M2) and (M3) are

CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein

XX SQ Sequence 395 AA;

Query Match 45.4%; Score 1265.5; DB 6; Length 395;
Best Local Similarity 65.2%; Pred. No. 1.9e-68;
Matches 255; Conservative 59; Mismatches 50; Indels 27; Gaps 5;

QY	169	GKEITLIMQTLNTLSTPEEKLAALCKKYAELLBEHNSOKMGLQKKQSQLYQEKDHLR	228
Db	3	GKEVLIMQALNTLSTPEEKLAALCKKYADLLSESRVQMKILQKQAOIVKEKRVHLQ	62
QY	229	GEHSAVLAEKSLRLCRQLQRNRSIKKEGVORARERERERKEVTSHFQVTLNDIQLQM	288
Db	63	SEHSAVLAEKSLRLCRQLQRNRSIKKEGVORARERERERKEVTSHFQVTLNDIQLQM	122
QY	289	EQHNRNSKLRQRNRLAEKSLRLCRQLQRNRSIKKEGVORARERERERKEVTSHFQVTLNDIQLQM	348
Db	123	EQHNRNSKLRQRNRLAEKSLRLCRQLQRNRSIKKEGVORARERERERKEVTSHFQVTLNDIQLQM	182
QY	349	KEAEERHOREKDFLLKEAVESQRMCELMKQOETHLQOOLALYTKFEEFQNTLSKSEVF	408
Db	183	KEAEERHOREKDFLLKEAVESQRMCELMKQOETHLQOOLALYTKFEEFQNTLSKSEVF	242
QY	409	TTFRQEMEMTKKIKKLEKETTWYRSRWSSNKLLEMAEETKTVRDKELBGLQVIOIRLE	468
Db	243	TTFRQEMEMTKKIKKLEKETTWYRSRWSSNKLLEMAEETKTVRDKELBGLQVIOIRLE	302
QY	469	KLCRALQTERNDLKNRVQ-----DL-SAGGQGLT--DSGPE-----RR	504
Db	303	KLCRALQTERNDLKNRVQ-----DL-SAGGQGLT--DSGPE-----RR	362
QY	505	PEGPGAQA-PSSPR--VTEAPCYFGAPSTEA	532
Db	363	ALGAHLEAEPKYSQSAVQKPPSTGSAPAIES	393

Search completed: June 8, 2004, 16:32:54
Job time : 62 secs

QY 241 LESLCRLQHRNRSLEKGGVQRAEERKKEVTSHTFQVTLNDIQLQMEQHNRNRSKLQ 300
DB 241 LESLCRLQHRNRSLEKGGVQRAEERKKEVTSHTFQVTLNDIQLQMEQHNRNRSKLQ 300
QY 301 ENMELAEKLLKLIQYELREBHIDKVFHKLQQLVDKLLQQLVDKLLQQAQEMLEAEERHOREKD 360
DB 301 ENMELAEKLLKLIQYELREBHIDKVFHKLQQLVDKLLQQLVDKLLQQAQEMLEAEERHOREKD 360
QY 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKEFEFQNTLSKSSEVFTTFKQMEKMTX 420
DB 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKEFEFQNTLSKSSEVFTTFKQMEKMTX 420
QY 421 KIKLEKETTMYRNRWESSNKALLLEAEKTVRKLEGLQVKIQRLEKLCRALQTERND 480
DB 421 KIKLEKETTMYRNRWESSNKALLLEAEKTVRKLEGLQVKIQRLEKLCRALQTERND 480
QY 481 LNKRVQDLASQGGSLTDSGPRRPEGGAQAPSSPRVTEAPCPYGPAPSTASQGTGPOE 540
DB 481 LNKRVQDLASQGGSLTDSGPRRPEGGAQAPSSPRVTEAPCPYGPAPSTASQGTGPOE 540
QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 2
US-08-979-608A-8
; Sequence 8, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arizona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV. 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-08-979-608A-8
Query Match 96.7%; Score 2693; DB 4; Length 530;
Best Local Similarity 99.8%; Pred. No. 1.6e-172; Indels 0; Gaps 0;
Matches 529; Conservative 0; Mismatches 1;
QY 17 KSPGQPEAGPEGAQERPQSAAPAVEAEGFGSSQAPRKEGAQARTAQSGALRDVSEELS 76
DB 1 KSPGQPEAGPEGAQERPQSAAPAVEAEGFGSSQAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVDNNGGGEAGGAPPAEDAEKSRTYVARNGEPETPTVVVNGKEPS 136
DB 61 RQLEDILSTYCVDNNGGGEAGGAPPAEDAEKSRTYVARNGEPETPTVVVNGKEPS 120
QY 137 KGPNTTEEIRQSDDEVDRHRPQEKKAAGLKGKJITLLMQTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTTEEIRQSDDEVDRHRPQEKKAAGLKGKJITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELEEHRSQKMLLQKQSLVQEKDHLRGEHSKAVLARSKLESICRELQHRNRSLK 256
DB 181 AELEEHRSQKMLLQKQSLVQEKDHLRGEHSKAVLARSKLESICRELQHRNRSLK 240
QY 257 EGVQARAREEEKKEVTSHTFQVTLNDIQLQMEQHNRNRSKLQENMELAEKLLKLIQY 316
DB 241 EGVQARAREEEKKEVTSHTFQVTLNDIQLQMEQHNRNRSKLQENMELAEKLLKLIQY 300
QY 317 ELREEHIDKVFHKLQQLVDKLLQQAQEMLEAEERHOREKD FLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVFHKLQQLVDKLLQQAQEMLEAEERHOREKD FLLKEAVESQRMCELM 360
QY 377 KQOETHLKQALYTEKEFEFQNTLSKSSEVFTTFKQMEKMTXKIKLEKETTMYRSRW 436
DB 361 KQOETHLKQALYTEKEFEFQNTLSKSSEVFTTFKQMEKMTXKIKLEKETTMYRSRW 420
QY 437 ESSNKALLEAEKTVRKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLASQGGSL 496
DB 421 ESSNKALLEAEKTVRKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLASQGGSL 480
QY 497 TDSGPRRPEGGAQAPSSPRVTEAPCPYGPAPSTASQGTGPOEPTSARA 546
DB 481 TDSGPRRPEGGAQAPSSPRVTEAPCPYGPAPSTASQGTGPOEPTSARA 530
RESULT 3
US-09-517-849-8
; Sequence 8, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arizona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608

FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-517-849-8

Query Match
Best Local Similarity 96.7%; Score 2693; DB 4; Length 530;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 76
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 60
QY 77 ROLEDILSTYCVDNNOGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 ROLEDILSTYCVDNNOGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KGPNTETIRQSDVGDHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTETIRQSDVGDHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 180
QY 197 AELLBEHNSQOMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLESCLRELQRNRSK 256
DB 181 AELLBEHNSQOMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLESCLRELQRNRSK 240
QY 257 EGVQARAEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 316
DB 241 EGVQARAEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
QY 317 ELREHIDKVPKHDLQQLQVDAKLQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREHIDKVPKHDLQQLQVDAKLQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQETHLQKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKETTMYRSW 436
DB 361 KQETHLQKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKETTMYRSW 420
QY 437 ESSNKALLEMAEKTVRDKEGLEQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLL 496
DB 421 ESSNKALLEMAEKTVRDKEGLEQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLL 480
QY 497 TDGSPERRPGEQAQAPSSPRVTBAPCYPGAPSTEASGQTGPQEPPTSARA 546
DB 481 TDGSPERRPGEQAQAPSSPRVTBAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 4
US-09-516-289-8
Sequence 8, Application US/09616289
Patent No. 6632923
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
TITLE OF INVENTION: ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/616,289
CURRENT FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 530
TYPE: PRT
ORGANISM: Homo sapiens
US-09-616-289-8

Query Match
Best Local Similarity 96.7%; Score 2693; DB 4; Length 530;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 76
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 60
QY 77 ROLEDILSTYCVDNNOGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 ROLEDILSTYCVDNNOGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KGPNTETIRQSDVGDHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTETIRQSDVGDHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 180
QY 197 AELLBEHNSQOMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLESCLRELQRNRSK 256
DB 181 AELLBEHNSQOMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLESCLRELQRNRSK 240
QY 257 EGVQARAEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 316
DB 241 EGVQARAEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
QY 317 ELREHIDKVPKHDLQQLQVDAKLQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREHIDKVPKHDLQQLQVDAKLQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQETHLQKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKETTMYRSW 436
DB 361 KQETHLQKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKETTMYRSW 420
QY 437 ESSNKALLEMAEKTVRDKEGLEQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLL 496
DB 421 ESSNKALLEMAEKTVRDKEGLEQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLL 480
QY 497 TDGSPERRPGEQAQAPSSPRVTBAPCYPGAPSTEASGQTGPQEPPTSARA 546
DB 481 TDGSPERRPGEQAQAPSSPRVTBAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 5
US-08-979-608A-5
Sequence 5, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street

```

QY 529 STEASGQTGPQBPTSARA 546
DB 540 STEASGQTGPQBETTATA 557

RESULT 6
US-09-517-849-5
; Sequence 5, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-Nov-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-517-849-5

Query Match 91.7%; Score 2552.5; DB 4; Length 557;
Best Local Similarity 91.0%; Pred. No. 4.4e-163;
Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

QY 1 MQNQDKKNGAAROSNPKSPGCGPEAGPEGAGQEPSPGAAAPAAVEAGPGSSQAPRKPEGAQA 60
DB 1 MQNQDKKNGAARQNPSSPGCGPEAGAEAGQGPAPAPAPAEAG-ASSAPAPGPEGAQA 59
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVDNNGGPGEDGAQGPPAPEDAEKGRITYVARN 120
DB 60 KTAQPGALCDVSELSRQLEDILSTYCVDNNGAPGEDGVQGPPEPDEAEKSRAYVARN 119
QY 121 GEPEP-TPVVGSEKYSKGDPTETEEIRQSDVGDGRDHRPQEKKAKGLCKEITLQMOTL 179
DB 120 GEPEPGTPVVGSEKETSKEAPGTETERTSDVGDGRDHRPQEKKAKGLCKEITLQMOTL 179
QY 180 NTLSTPEKLAALCKKYAEILLEHRNSQOMKLLQKQSOLVQEKDHLREHSAVLARS 239
DB 180 NTLSTPEKLAALCKKYAEILLEHRNSQOMKLLQKQSOLVQEKDHLREHSAVLARS 239

```

QY 180 NTLSTPEERKLAALCKKYAELLBEHRNSOKOMKLLQKQSOVLQOEKHLRGEHSKAVLARS 239
DB 180 NTLSTPEERKLAALCKKYAELLBEHRNSOKOMKLLQKQSOVLQOEKHLRGEHSKAVLARS 239
QY 240 KLESICRELOHNRSLKEEGVQVORAREEKEKVTSHFQVTLNDIOLQOEQHNRNSKLR 299
DB 240 KLESICRELOHNRSLKEEGVQVORAREEKEKVTSHFQVTLNDIOLQOEQHNRNSKLR 299
QY 300 QENWELAEKRLKLEIQYELREEHIDKVFHKLQOQOLVDAKLAQQAQEMLKEAEERHOREK 359
DB 300 QENWELAEKRLKLEIQYELREEHIDKVFHKLQOQOLVDAKLAQQAQEMLKEAEERHOREK 359
QY 360 DFLKKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQNTLSKSEVFTTFQEMERMT 419
DB 360 DFLKKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQNTLSKSEVFTTFQEMERMT 419
QY 420 KIKKLEKETTYRSTRWESSNKALLEMAEKTVDKLEGLQVVKIQRLEKLCALQOTERN 479
DB 420 KIKKLEKETTYRSTRWESSNKALLEMAEKTVDKLEGLQVVKIQRLEKLCALQOTERN 479
QY 480 DLNKRVDLSAGGQSLTDSGPRRP-----EGFGAQPSSPRVTAPCPYPGAP 528
DB 480 DLNKRVDLSAGGQSLTDSGPRRP-----EGFGAQPSSPRVTAPCPYPGAP 528
QY 529 STEASGOTGPQEPPTATA 546
DB 540 STEASGOTGPQEPPTATA 557

RESULT 8

US-08-968-751-6
; Sequence 6, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfeld, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligandfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/968,751
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Giotta, Gregory
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX1024 GG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 262-8710
; TELEFAX: (510) 222-9758
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-968-751-6

Query Match 37.3%; Score 1039.5; DB 2; Length 386;
Best Local Similarity 56.0%; Pred. No. 4.8e-62;

QY 240 KLESICRELOHNRSLKEEGVQVORAREEKEKVTSHFQVTLNDIOLQOEQHNRNSKLR 299
DB 240 KLESICRELOHNRSLKEEGVQVORAREEKEKVTSHFQVTLNDIOLQOEQHNRNSKLR 299
QY 300 QENWELAEKRLKLEIQYELREEHIDKVFHKLQOQOLVDAKLAQQAQEMLKEAEERHOREK 359
DB 300 QENWELAEKRLKLEIQYELREEHIDKVFHKLQOQOLVDAKLAQQAQEMLKEAEERHOREK 359
QY 360 DFLKKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQNTLSKSEVFTTFQEMERMT 419
DB 360 DFLKKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQNTLSKSEVFTTFQEMERMT 419
QY 420 KIKKLEKETTYRSTRWESSNKALLEMAEKTVDKLEGLQVVKIQRLEKLCALQOTERN 479
DB 420 KIKKLEKETTYRSTRWESSNKALLEMAEKTVDKLEGLQVVKIQRLEKLCALQOTERN 479
QY 480 DLNKRVDLSAGGQSLTDSGPRRP-----EGFGAQPSSPRVTAPCPYPGAP 528
DB 480 DLNKRVDLSAGGQSLTDSGPRRP-----EGFGAQPSSPRVTAPCPYPGAP 528
QY 529 STEASGOTGPQEPPTATA 546
DB 540 STEASGOTGPQEPPTATA 557

RESULT 7

US-09-616-289-5
; Sequence 5, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-616-289-5

Query Match 91.7%; Score 2552.5; DB 4; Length 557;
Best Local Similarity 91.0%; Pred. No. 4.4e-183;
Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;
QY 1 MKNQDKNGAAGKQNPSSPCQPEAGPEGAQPSQAAPAEAGPGSSQAPRKEGAQA 60
DB 1 MKNQDKNGAAGKQNPSSPCQPEAGPEGAQPSQAAPAEAGPGSSQAPRKEGAQA 59
QY 61 RTAQSGALRDVSELSQLEIDILSTYCVDNQGGPDGAGQGPAPAEPAEDAEKRTYVARN 120
DB 60 RTAQSGALRDVSELSQLEIDILSTYCVDNQGGPDGAGQGPAPAEPAEDAEKRTYVARN 119
QY 121 GEPEP-TPVNVGRKEPKSGDNTPEIRQSDVGVGDHRRPOEKKKAGLGKETTLLMOTL 179
DB 120 GEPEP-TPVNVGRKEPKSGDNTPEIRQSDVGVGDHRRPOEKKKAGLGKETTLLMOTL 179

Matches 225; Conservative 55; Mismatches 99; Indels 23; Gaps 6;
QY 29 QAQRPQQAAPAVEAGPGSSQAARK--PEGAQARTAQSGALRDVSBELSGLEDILSTY 86
DB 4 GAEE-----ATEAGRGRRRSRQKPEIGTMEBAGICGLGVXADMCLNSQSNLILQHQ 56
QY 87 CVDNNOGPGEDGAGGAPAEADAESRTYVARNGEPPTPVNGEKPSKGDPTTEIR 146
DB 57 --GSCGTSNKHSLSEDEGSDPTENNLVS-----PAYCTOESREIEPG--EAR 104
QY 147 QSDVGDHRDRRPOEKXKARGIKREITLLMQLNTLSTPEKLAALCKKYAELLEHRNS 206
DB 105 TDPDQDQSECNRNKEKT--LGKVLLEMLQALNTLSTPEKLAALCKKYADLLEBSRSV 162
QY 207 QKQMKLQKQSOVLQOEKHLRGHESKAVLARSKLESICRLQHRNLSKESGVQARER 266
DB 163 QKQMKLQKQQAQIVKEKVLQSESKAILARSKLESICRLQHRNLSKESGVQARER 222
QY 267 EEKKEVTSHFQVTLNDIQLQMEQCHNRSNKLQRMELAEKLLIEQYELREHIDKV 326
DB 223 EERRTEATAHQITLNEIQALQEQHDHNAKLQENIELGEKLLIEQYALREHIDKV 282
QY 327 FKHKDLOOLVDKLOQAOEMLKEABERHOREKDFLLKEAVESQRMCELMKQETHLQO 386
DB 283 FKHKDLOOLVDKLOQAOEMLKEABERHOREKDFLLKEAVESQRMCELMKQETHLQO 342
QY 387 LALYTEKTEEFONTLSKSEVFTTTPKQEMEKMTKKIKLEKE 428
DB 343 LSLYMDKKEEFTTAKSNELFTTTPKQEMEKMTKKIKKXKK 384
RESULT 9
US-09-688-188B-151
; Sequence 151, Application US/09688188B
; Patent No. 6656716
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0328
; CURRENT APPLICATION NUMBER: US/09/688,188B
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 09/291,417
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: 60/081,784
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-188B-151
Query Match 9.0%; Score 252; DB 4; Length 787;
Best Local Similarity 21.1%; Pred. No. 5.7e-09;
Matches 140; Conservative 121; Mismatches 215; Indels 186; Gaps 29;
QY 4 QDKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQA-PRKPEGAQART 62
DB 104 QEAHNGPAYEAQEEEEEQDHGVGRTGT-----VNSVGSNQSIPMSISASSQS 151
QY 63 AQSGALRDVSBELSGLEDILSTYCVDDNQG---GPGEDG--AQGEPAEPEDAESRTY 116
DB 152 SSVNSLPDVSDKKS-ELDMMEGDHTVMSNSVTHLKPEENYREGDP-----RT- 200
QY 117 VARNGEPPTPVVNGEKPSKGDPTTEIRQSDVGDHRDRRPOEKKAKGLGKAIT--- 173
DB 201 --RASDPQGPQVSRHKSHYRNREHATIRTASLV---TROMQEHQDSLEQMSGYK 254
QY 174 -LLMQLNTLSTPEKLAALCKKYA---ELLEHEN--SQKMKLQKQSQCLQVE--- 223

DB 255 RMRQHQKQLMTLENKLAEMDEHRLRLDKCLTETGRNNPAAEMEXLKKHQAAMEKAV 314
QY 224 -----KDHLEGEHSAVLARSKLESICRELQHRNLSKESGVOR-----AREEE 268
DB 315 MSNEEKFKQHQIAQOQKEL--NSFLESQKREYKLRKLEQLKEELNENQSTPKKQEWLS 372
QY 269 KRKEVTSHFQV--TLNDIOLQME-----OHN-----ERNSKLRQEN 302
DB 373 KQENLQHQQAEBEANLLRQRYLELCRPFKRMMLGRENLEQDLVREELNKRQTKD 432
QY 303 MELAEKLLIEQYELREHIDKVFKHKDLOOLVDKLOQAOEMLKEABERHOREKDFL 362
DB 433 LEHALLRQHESMQELFEHRLNTIQK---MRCELI--RLQHQTELNTQLEYNKRERELR 487
QY 363 LKEAVESQRMCELMKQETHLQOQ---ALYTEKTEEFONTLSKSEVFTTTPKQEMEK 418
DB 488 RKHVEVRQPKSLKSEKLIQKQFQDTCKIQTRQYKALRNHLE-----TTPKSEHKA 542
QY 419 TKKIKKLEKETMYRWRBESSNKALLEMAEKTVR-----453
DB 543 LKRLK--EQTRKLAILEAQYDHSINEMLSLSTQALRLDEAQAECQVLMQLOQELBLNA 600
QY 454 -----DKLEGLQVKI-----ORLEKLCRALCTERND-----LNKRQVD 487
DB 601 YQSKIQMAQERQHDRELRELEQRVSLRALLEQKIIEMLALQNERTERIRSLERQARE 660
QY 488 LSAQGGQS-----LTDSGFERPPEGPGAQAPSPRVTEAPCTYPGAP--STEASQOT 536
DB 661 IEAPDSSEMLGFSNMVLSNLSPE-----AFSHSYPGASGWSHNPCTGPP 704
QY 537 GP 538
DB 705 GP 706
RESULT 10
US-09-291-417D-151
; Sequence 151, Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/081,784
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-151
Query Match 9.0%; Score 252; DB 4; Length 787;
Best Local Similarity 21.1%; Pred. No. 5.7e-09;
Matches 140; Conservative 121; Mismatches 215; Indels 186; Gaps 29;
QY 4 QDKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQA-PRKPEGAQART 62
DB 104 QEAHNGPAYEAQEEEEEQDHGVGRTGT-----VNSVGSNQSIPMSISASSQS 151
QY 63 AQSGALRDVSBELSGLEDILSTYCVDDNQG---GPGEDG--AQGEPAEPEDAESRTY 116
DB 152 SSVNSLPDVSDKKS-ELDMMEGDHTVMSNSVTHLKPEENYREGDP-----RT- 200
QY 117 VARNGEPPTPVVNGEKPSKGDPTTEIRQSDVGDHRDRRPOEKKAKGLGKAIT--- 173
DB 201 --RASDPQGPQVSRHKSHYRNREHATIRTASLV---TROMQEHQDSLEQMSGYK 254

QY 174 -LLMOTLNTLSTPEEKLAALCKKYA---ELLEHN--SOKMKLLQKQSOLOV--- 223
DB 255 RMRQHCKQLMTLENLKAEMDEHRLRLDKLETGNNFAAEMEKLIKQQAAMEKEAV 314
QY 224 -----KDHRLGEHSAVLARSKLESICRELORHRSLEKEGVOR-----ARBEER--- 268
DB 315 NGNEEXXFOOHIQAQCKKEL--NSFLESQKREYKLRKQOLKBELENNQSTPKKEKQEWLS 372
QY 269 KRKEVTSHPQV--TLNDIOLQME-----QHN-----ERNKSLROEN 302
DB 373 KQKENIQHQAEBEANLRRQYLECECRPKRMLGRHNLQDLVREELNKRQTKD 432
QY 303 MELAEKLLIBOYELREHIDKVFHKLQLOOLVDKLOQAQEMKBAEERHOREKDFL 362
DB 433 LEHAMLARQHSQOELEFRHLATIQ--MRCELI--RLQHOTELTNQLEYNKRERELR 487
QY 363 LKEAVESORWELMKQOETHLKOOL---ALYTEKEPEFQNTLSKSEVFTTFKQEMECM 418
DB 488 RHMVMEVRQPSKLSKELQIKQFODTKCIQTRQYKALRNHLE-----TTPKSEHKA 542
QY 419 TKKIKKLEKETTMYRSWSSNKALLEMAEKTVR----- 453
DB 543 LKRLK--EETRKLAIALAEQYDHSINEMLSLQALRLDEAQAECQVLKQMLQOELELINA 600
QY 454 -----DKELEGLOVKI-----ORLEKLCRALQTERND-----LNKRVOD 487
DB 601 YQSKIQAQAQHDREHLEQVLSRRALLEGKIEEMMALQNERTERIRSLERQARE 660
QY 488 LSAGGQGS-----LTDGSPERPPEGGAQAPSSPRVTEAPCYGAP--STASGOT 536
DB 661 IBAPDSMRLGFSNMVLSNLSPE-----AFSHSYPGASGSHNFTGPGP 704
QY 537 GP 538
DB 705 GP 706

RESULT 11
US-09-688-188B-23
; Sequence 23, Application US/09688188B
; Patent No. 6656716
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0328
; CURRENT APPLICATION NUMBER: US/09/688,188B
; PRIOR FILING DATE: 2000-10-16
; PRIOR FILING DATE: 09/291,417
; PRIOR FILING DATE: 1999-04-14
; PRIOR FILING DATE: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-188B-23

Query Match 9.0%; Score 251.5; DB 4; Length 786;
Best Local Similarity 21.6%; Pred. No. 6.1e-09;
Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;
QY 39 PAVANE-----GPGSSQA-PRKEGAQAARTAQSGALRVSELSRQLE 80
DB 109 PAVAQEEBEEQDHGVRTVNSVGNQSIPEMSISASQSSVNSLPVSDKX-ELD 167
QY 81 DILSTYCVDNNGQ---GPGEDG---AQGEPAEPEDAERSRTYVARNGEPEPTPVNKEKE 134
DB 168 MMEGDHTVMSNSVHLKPEENYREBGP-----RT---RASDPQSPQVSRHKS 215

QY 135 PSKGDPTTEIROQSDVGDRRHRRPOEKKKAGLGKEIT-----LLMOTLNTLSTPEEKLA 190
DB 216 HYRREHPATIRASLV---TRQWEEHQDSERLEQMSGYKRMERORQKQMLTLENLKL 271
QY 191 ALCKKYA---ELLEHN--SOKMKLLQKQSOLOV---KDHRLGEHSA 233
DB 272 AKMEHRLRLDKLETGNNFAAEMEKLIKQQAAMEKEAVNSNEEKFOOHIQAQCKK 331
QY 234 AVLARSKLESICRELORHRSLEKEGVOR-----ARBEER---KRKEVTSHPQV--TLNDI 284
DB 332 EL--NSFLESQKREYKLRKQOLKBELENNQSTPKKEKQEWLSKQKENTQHPQAEBEANLL 389
QY 285 OLQME-----QHN-----ERNKSLROENMELABRLKLLIEQYELRE 320
DB 336 RRQRYLECECRPKRMLGRHNLQDLVREELNKRQTKDLEHAMLARQHSQOELEF 449
QY 321 EHIKRVFHKDLQLOOLVDKLOQAQEMKBAEERHOREKDFLKEAVESORWELMKQOE 380
DB 450 RLHNTIQ--MRCELI--RLQHOTELTNQLEYNKRERELRKHVMEVRQPSKLSKSE 504
QY 381 THLKOOL---ALYTEKEPEFQNTLSKSEVFTTFKQEMEKTKIKLEKETMYRSRW 436
DB 505 LQIKKQFODTKCIQTRQYKALRNHLE-----TTPKSEHKAHLKEL--EETRKLAIALA 557
QY 437 ESSNKALLEMAEKTVR----- 458
DB 558 EYDHSINEMLSLQALRLDEAQAECQVLKQMLQOELELINAQSKIRMQAQAQHDRELR 617
QY 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGQGS-----L 496
DB 618 ELEQVLSRRALLEGKIEEMMALQNERTERIRSLERQAREIEAFDSMRLGFSNMVL 677
QY 497 TDSGPERPPEGGAQAPSSPRVTEAPCYGAP--STASGOTGP 538
DB 678 SNLSPE-----AFSHSYPGASGSHNFTGPGP 705

RESULT 12
US-09-291-417D-23
; Sequence 23, Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13
; PRIOR FILING DATE: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-23

Query Match 9.0%; Score 251.5; DB 4; Length 786;
Best Local Similarity 21.6%; Pred. No. 6.1e-09;
Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;
QY 39 PAVANE-----GPGSSQA-PRKEGAQAARTAQSGALRVSELSRQLE 80
DB 109 PAVAQEEBEEQDHGVRTVNSVGNQSIPEMSISASQSSVNSLPVSDKX-ELD 167
QY 81 DILSTYCVDNNGQ---GPGEDG---AQGEPAEPEDAERSRTYVARNGEPEPTPVNKEKE 134
DB 168 MMEGDHTVMSNSVHLKPEENYREBGP-----RT---RASDPQSPQVSRHKS 215
QY 135 PSKGDPTTEIROQSDVGDRRHRRPOEKKKAGLGKEIT-----LLMOTLNTLSTPEEKLA 190

Db 216 HYNREHFAITIRASIV---TRQMEHQDSELREQMSGYKMRQHQKQALATLENK 271
 QY 191 ALCKKYA---ELLBEHRN--SOKOMKLOKQOSQVQV-----KDHURGEHSK 233
 Db 272 ARMDERLRDLKLETORNFNFAEMKLIKQQAAMEKAQWMSNEEKFFQCHIQAOQKK 331
 QY 234 AVLARSKLESCLRELQRNRSLSKEGVQR---AREBER---KKEVTSHFOV--TLNDI 284
 Db 332 EL--NSFLESQREYKRLKEQLKEHLENQSTPKKEKQEWLSKQENIQHQAABEENLL 389
 QY 285 QLOME-----QHN-----ERNSKLRQENMELARLKKLIROYELRE 320
 Db 390 RRQRYLELCRRFRKRMMLGRHLEQDLVREELNKQTKQDLEHMLLRQHESMOBLEF 449
 QY 321 EHDVKVFKHDLQOOLVDKLAQQAQEMLKEAERHOREKDFLLKEAVSQRMCELMKQOE 380
 Db 450 RHNTIQK---MRCELI--RLQHOTELTNQLEYNKRREBLRRKHVMVEVROQPSLSKSE 504
 QY 381 THLKOOL-----ALYTEKPEFQNTLSKSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436
 Db 505 LQIKKQFQDTCKIQTRYKALRNHLL-----TTPKSEHKAVALKRLK--EQTRKLAILA 557
 QY 437 ESSNKALLEMAEKTVR-----DKELE 458
 Db 558 EYDHSINEMLSQALRLDEAQAECQVLKMLQLOQLELLNAYQSKIKMQAQAQHDREL 617
 QY 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVQDLSAGGGS-----L 496
 Db 618 ELEQVSLRRALLEQKIBEMALQNERIRSLERQAREIAFDSSEYELGFSNNVL 677
 QY 497 TDSGPRRPEGGAQAPSPRVTEAPCYCAP--STEASGQTGP 538
 Db 678 SNLSPE-----APSHSYPGASGWSHNPTGGPGP 705

RESULT 13

US-09-688-1888-31
 ; Sequence 31, Application US/096881888
 ; Patent No. 6656716

; GENERAL INFORMATION:
 ; APPLICANT: FLOWMAN, GREGORY
 ; APPLICANT: MARTINEZ, RICARDO
 ; APPLICANT: WHYTE, DAVID
 ; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
 ; FILE REFERENCE: 038602/0328
 ; CURRENT APPLICATION NUMBER: US/09/688,188B
 ; CURRENT FILING DATE: 2000-10-16
 ; PRIOR APPLICATION NUMBER: 09/291,417
 ; PRIOR FILING DATE: 1999-04-14
 ; PRIOR APPLICATION NUMBER: 60/081,784
 ; PRIOR FILING DATE: 1998-04-14
 ; NUMBER OF SEQ ID NOS: 155
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 31
 ; LENGTH: 1001
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: murine/human SULU3
 US-09-688-1888-31

Query Match 9.0%; Score 251.5; DB 4; Length 1001;
 Best Local Similarity 21.6%; Pred. No. 8.3e-09;
 Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;
 QY 39 PAVEAB-----GPOSSQA-PRKEGQAQARTAGSALRDVSELSRQLE 80
 Db 324 PAVEAQEEEBEODHGVGRGTGTVNSVGSNQSIPTSMISASSQSSVNSLDPVSDDKS-ELD 382
 QY 81 DILSTYCVNNQO---GPEDEG--AQCEPAEPEDAERKRTYVARGEPPTPVNGEKE 134
 Db 383 MMEGDHTVMSNSVTHLKEENYREGDP-----RT---RASDPQSPQVSRHKS 430

QY 135 PSKGDPTNEITROSDEYGRDRHRRPOBKAKKAGLGKEIT---LLMOTLNTLSTPEKLA 190
 Db 431 HYNREHFAITIRASIV---TRQMEHQDSELREQMSGYKMRQHQKQALATLENK 486
 QY 191 ALCKKYA---ELLBEHRN--SOKOMKLOKQOSQVQV-----KDHURGEHSK 233
 Db 487 ARMDERLRDLKLETORNFNFAEMKLIKQQAAMEKAQWMSNEEKFFQCHIQAOQKK 546
 QY 234 AVLARSKLESCLRELQRNRSLSKEGVQR---AREBER---KKEVTSHFOV--TLNDI 284
 Db 547 EL--NSFLESQREYKRLKEQLKEHLENQSTPKKEKQEWLSKQENIQHQAABEENLL 604
 QY 285 QLOME-----QHN-----ERNSKLRQENMELARLKKLIROYELRE 320
 Db 605 RRQRYLELCRRFRKRMMLGRHLEQDLVREELNKQTKQDLEHMLLRQHESMOBLEF 664
 QY 321 EHDVKVFKHDLQOOLVDKLAQQAQEMLKEAERHOREKDFLLKEAVSQRMCELMKQOE 380
 Db 665 RHNTIQK---MRCELI--RLQHOTELTNQLEYNKRREBLRRKHVMVEVROQPSLSKSE 719
 QY 381 THLKOOL-----ALYTEKPEFQNTLSKSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436
 Db 720 LQIKKQFQDTCKIQTRYKALRNHLL-----TTPKSEHKAVALKRLK--EQTRKLAILA 772
 QY 437 ESSNKALLEMAEKTVR-----DKELE 458
 Db 773 EYDHSINEMLSQALRLDEAQAECQVLKMLQLOQLELLNAYQSKIKMQAQAQHDREL 832
 QY 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVQDLSAGGGS-----L 496
 Db 833 ELEQVSLRRALLEQKIBEMALQNERIRSLERQAREIAFDSSEYELGFSNNVL 892
 QY 497 TDSGPRRPEGGAQAPSPRVTEAPCYCAP--STEASGQTGP 538
 Db 893 SNLSPE-----APSHSYPGASGWSHNPTGGPGP 920

RESULT 14

US-09-291-417D-31
 ; Sequence 31, Application US/09291417D
 ; Patent No. 6680170

; GENERAL INFORMATION:
 ; APPLICANT: FLOWMAN, GREGORY
 ; APPLICANT: MARTINEZ, RICARDO
 ; APPLICANT: WHYTE, DAVID
 ; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
 ; FILE REFERENCE: 038602/0329
 ; CURRENT APPLICATION NUMBER: US/09/291,417D
 ; CURRENT FILING DATE: 1999-04-13
 ; PRIOR APPLICATION NUMBER: 60/081,784
 ; PRIOR FILING DATE: 1998-04-14
 ; NUMBER OF SEQ ID NOS: 155
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 31
 ; LENGTH: 1001
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: murine/human SULU3
 US-09-291-417D-31

Query Match 9.0%; Score 251.5; DB 4; Length 1001;
 Best Local Similarity 21.6%; Pred. No. 8.3e-09;
 Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;
 QY 39 PAVEAB-----GPOSSQA-PRKEGQAQARTAGSALRDVSELSRQLE 80
 Db 324 PAVEAQEEEBEODHGVGRGTGTVNSVGSNQSIPTSMISASSQSSVNSLDPVSDDKS-ELD 382
 QY 81 DILSTYCVNNQO---GPEDEG--AQCEPAEPEDAERKRTYVARGEPPTPVNGEKE 134


```

Db 383 MMEGDTVMSSSVIHLKPEENYREGDP-----RT---RASDPQSPPOVSRHKS 430
Qy 135 PSKGDPTTEIRQSDVEGDRHRRPOEKKAGLGKEIT-----LLMQTLNTLSTPEZKLA 190
Db 431 HYRNEHPATITASLV-----TRQOEHEQDSELRQMSGYRMRHQKQOLATLENKLG 486
Qy 191 ALCKKYA-----ELLEHRN--SQOMKLLQKQSQVQ3-----KDLHRLGBHSHK 233
Db 487 AKMDHRLRLDKLQRTNNFAAEKELIKKQAAEKEAKVMSNEEKKFQOHLQAOQKK 546
Qy 234 AVLARSKLESICRELQRNRSLSKEGVQV-----AREEE--KRKEVTSHFQV--TLNDI 284
Db 547 EL--NSFLESQKREYKLRKQBLKEBLNENQSTPKKEQEWLSKQENIQHFQAEERANLL 604
Qy 285 QLOME-----QHN-----ERNKSLRQENMLAEKRLKGLIEOYELRE 320
Db 605 RRQOYLECRRFRKRMMLGRHNLQDLVREELNKRQTKDLEHMLLRQHSQOHELP 664
Qy 321 EHDKVFHKLQOOLVDAKLQAOEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOE 380
Db 665 RHLNTIQK---MRCELI--RLQHQTETLNQLEYNKRERERELRRKHVMEVROQPKSLAKSE 719
Qy 381 THLKQOL---ALYTEKPEEFONTLSKSSEVFTTFKQEMKEMTKIKKLEKETTMYRSRW 436
Db 720 LQIKKQFDTCIKITQRYKALRNHLE-----TPKSEHKAVLKLK--EQTRKLAILAQ 774
Qy 437 ESSNKALLEMAEKTVR-----DRELE 458
Db 773 EGYDSINEMLSQALRLDEAQAECQVLKMQLOQLELLNAYQSKIKNQAEQAQHDREL 832
Qy 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS 496
Db 833 ELEQVSLRALLQKIEEMALQNERTERIRSLERQAREIEAFDSMESMLGFSNNVL 892
Qy 497 TDSGPERRPEGAQAPS-SPRVTEAP-CYPGAP--STEASGQTGP 538
Db 893 SNLSPE-----APSHSYGASGSHNFTGPGP 920

```

RESULT 15

```

US-09-060-410-2
; Sequence 2, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

```

```

; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-060-410-2

```

Query Match 9.0%; Score 250; DB 3; Length 1001;

```

Best local Similarity 21.9%; Pred. No. le-08;
Matches 142; Conservative 119; Mismatches 211; Indels 176; Gaps 30;

```

```

Qy 39 PAYEAE-----GPGSSQA-PRKPEGAOARTAQSGALRVSVSELSQLE 80
Db 324 PAVEAQEEBEEQDGGGRTCTVNSVGSNQSIPIWSISASSQSSVNSLPPASDDKS-ELD 382
Qy 81 DIUSTVCVDNQG-----GPGEDGAQGEPAEPEDAERKSTVYVARGEPETPTVNGEKEPS 136
Db 383 MMEGDTVMSSSVIHLKPEENYQ--EEGDP-----TRASAPQSPPOVSRHSHY 432
Qy 137 KGPNTTEIRQSDVEGDRHRRPOEKKAGLGKEIT-----LLMQTLNTLSTPEEKLAAL 192
Db 433 RNREHPATITASLV-----TRQOEHEQDSELRQMSGYRMRHQKQOLMTLENKAE 489
Qy 193 CKYA-----ELLEHRN--SQOMKLLQKQSQVQ3-----KDLHRLGBHSHKAV 235
Db 489 MDEHRLRLDKLQRTNNFAAEKELIKKQAAEKEAKVMSNEEKKFQOHLQAOQKKEL 548
Qy 236 LARSKLESICRELQRNRSLSKEGVQV-----AREEE--KRKEVTSHFQV--TLNDIOL 286
Db 549 --NSFLESQKREYKLRKQBLKEBLNENQSTPKKEQEWLSKQENIQHFQAEERANLRR 606
Qy 287 QME-----QHN-----ERNKSLRQENMLAEKRLKGLIEOYELREEH 322
Db 607 QROYLECRRFRKRMMLGRHNLQDLVREELNKRQTKDLEHMLLRQHSQOHELSFRH 666
Qy 323 IDKVFHKLQOOLVDAKLQAOEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETH 382
Db 667 LNTIQK---MRCELI--RLQHQTETLNQLEYNKRERERELRRKHVMEVROQPKSLAKELQ 721
Qy 383 LKQOL---ALYTEKPEEFONTLSKSSEVFTTFKQEMKEMTKIKKLEKETTMYRSRWES 438
Db 722 IKQFDQDTCIKITQRYKALRNHLE-----TPKSEHKAVLKLK--EQTRKLAILAQ 774
Qy 439 SNKALLEMAEKTVR-----DKELEGL 460
Db 775 YDHSINEMLSQALRLDEAQAECQVLKMQLOQLELLNAYQSKIKNQAEQAQHDRELREL 834
Qy 461 QVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS-----LTD 498
Db 835 EQVSLRALLQKIEEMALQNERTERIRSLERQAREIEAFDSMESMLGFSNNVLN 894
Qy 499 SGPERRPEGAQAPS-SPRVTEAP--CYPGAPSTEASG---QTPQ 539
Db 895 LSPEAFSHSYGASGSHNFTGPGHGHGHPMGGTPOAWGHPMGQGPQ 942

```

Search completed: June 8, 2004, 16:36:45

Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 8, 2004, 16:31:45 ; Search time 56 Seconds
(without alignments)
2743.052 Million cell updates/sec

Title: US-10-023-529-44
Perfect score: 2785
Sequence: 1 MKNQDKKNGAKQSNPKSSP.....APSTEASGQTGPQPTSA 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2785	100.0	546	9	US-09-976-740-44
2	2785	100.0	546	12	US-10-671-242-44
3	2785	100.0	546	13	US-10-023-529-44
4	2785	100.0	546	13	US-10-023-529-44
5	2785	100.0	546	15	US-10-616-187-44
6	2693	96.7	530	9	US-09-962-055-8
7	2693	96.7	530	9	US-09-976-740-8
8	2693	96.7	530	12	US-10-671-242-8
9	2693	96.7	530	13	US-10-023-529-8
10	2693	96.7	530	13	US-10-023-529-8
11	2693	96.7	530	15	US-10-616-187-8
12	2552.5	91.7	557	9	US-09-962-055-5
13	2552.5	91.7	557	9	US-09-976-740-5
14	2552.5	91.7	557	12	US-10-671-242-5
15	2552.5	91.7	557	13	US-10-023-529-5

16	2552.5	91.7	557	13	US-10-023-523-5
17	2552.5	91.7	557	13	US-10-616-187-5
18	2416	86.8	510	12	US-10-276-774-2134
19	501	16.9	437	12	US-10-425-114-55188
20	470.5	16.9	436	12	US-10-425-599-207434
21	444.5	16.0	314	12	US-10-425-114-43684
22	375	13.5	346	15	US-10-369-493-5130
23	304.5	10.9	206	12	US-10-425-114-45419
24	263	9.4	2871	14	US-10-146-473-41
25	260	9.3	886	15	US-10-369-493-1016
26	260	9.3	909	9	US-09-925-298-988
27	260	9.3	909	10	US-09-925-299-988
28	257.5	9.2	860	12	US-10-072-012-838
29	257.5	9.2	860	12	US-10-037-417-59
30	257.5	9.2	860	15	US-10-080-334-166
31	254.5	9.1	880	15	US-10-369-493-21643
32	253.5	9.1	1534	14	US-10-342-136-1
33	252.5	9.1	1881	14	US-10-032-585-7646
34	251.5	9.0	786	10	US-09-291-417-31
35	251.5	9.0	1001	10	US-09-291-417-31
36	251.5	9.0	1005	14	US-10-177-293-254
37	251.5	9.0	1005	14	US-10-317-835-18
38	250	9.0	1001	15	US-10-445-735-2
39	247	8.9	85	15	US-10-264-049-3413
40	244.5	8.8	3225	16	US-10-408-765A-254
41	240	8.6	650	15	US-10-104-047-3636
42	240	8.6	1001	14	US-10-128-714-3240
43	240	8.6	1162	12	US-09-894-273-2
44	240	8.6	1162	14	US-10-294-804-2
45	239	8.6	1137	12	US-10-336-472-16

ALIGNMENTS

RESULT 1

US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44

Query Match 100.0%; Score 2785; DB 9; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.2e-156;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKKNGAKQSNPKSSPGQPEAGPQERPSQAAPAVEAGPGSSQAPKPEGQA 60
DB 1 MKNQDKKNGAKQSNPKSSPGQPEAGPQERPSQAAPAVEAGPGSSQAPKPEGQA 60

;; PRIOR FILING DATE: 1997-11-26
;; PRIOR APPLICATION NUMBER: US 60/031,930
;; PRIOR FILING DATE: 1996-11-27
;; PRIOR APPLICATION NUMBER: US 60/048,547
;; PRIOR FILING DATE: 1997-06-03
;; NUMBER OF SEQ ID NOS: 53
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 44
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 100.0%; Score 2785; DB 15; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.2e-156;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKXGAKQNPSSPQPPGAGPEGAQERSQAAPVAEAGPGSSQAPRKEGAQA 60
DB 1 MNQDKXGAKQNPSSPQPPGAGPEGAQERSQAAPVAEAGPGSSQAPRKEGAQA 60
QY 61 RTAQSGLRDVSEBELSROLEDILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARN 120
DB 61 RTAQSGLRDVSEBELSROLEDILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARN 120
QY 121 GEPEPTPVVNGKEPSKDPNTEETIRQSDVGVDRDHRPQKKAKGLGKEITLLMOTLN 180
DB 121 GEPEPTPVVNGKEPSKDPNTEETIRQSDVGVDRDHRPQKKAKGLGKEITLLMOTLN 180
QY 181 TLSTPEEKLAALCKKYAELLBEHRNSQOMKLLQKQSQVQEKDHLRGHSHKAVLARSK 240
DB 181 TLSTPEEKLAALCKKYAELLBEHRNSQOMKLLQKQSQVQEKDHLRGHSHKAVLARSK 240
QY 241 LESLCELRQNRNSIKKEGVORAREBEERKEKVTSHFOVTNDIQLQMEQHNRNSKLQ 300
DB 241 LESLCELRQNRNSIKKEGVORAREBEERKEKVTSHFOVTNDIQLQMEQHNRNSKLQ 300
QY 301 ENMELAEKRLKXLEQYELREEHIDKVFHKDLQQLVDKLLQQAQEMLKEAEERHOREKD 360
DB 301 ENMELAEKRLKXLEQYELREEHIDKVFHKDLQQLVDKLLQQAQEMLKEAEERHOREKD 360
QY 361 FLLKEAVESORCELMKQOEHLKQALYTEKFEFQNTLSKSEVFTTFQEMEKMTK 420
DB 361 FLLKEAVESORCELMKQOEHLKQALYTEKFEFQNTLSKSEVFTTFQEMEKMTK 420
QY 421 KIKLEKETTYRSWESNKALEMABEKTVRDKEGLQVQIKLQLEKLCRALQTERND 480
DB 421 KIKLEKETTYRSWESNKALEMABEKTVRDKEGLQVQIKLQLEKLCRALQTERND 480
QY 481 LNKRVODLSAGQGSILTDGPPRRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQE 540
DB 481 LNKRVODLSAGQGSILTDGPPRRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQE 540
QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 6
US-09-962-055-8
; Sequence 8, Application US/09962055
; Patent No. US20020052033A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/962,055
FILING DATE: 24-Sep-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-962-055-8

Query Match 96.7%; Score 2693; DB 9; Length 530;
Best Local Similarity 99.8%; Pred. No. 3.1e-151;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGPEAGPEGAQRRPSQAAPVAEAGPGSSQAPRKEGAQAQTAQSGALRDVSEELS 76
DB 1 KSSPGPEAGPEGAQRRPSQAAPVAEAGPGSSQAPRKEGAQAQTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARNGBPEPTPVVNGKEPS 136
DB 61 RQLEDILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARNGBPEPTPVVNGKEPS 120
QY 137 KGPNTETIRQSDVGVDRDHRPQKKAKGLGKEITLLMOTNLSTPPEKLAALCKKY 196
DB 121 KGPNTETIRQSDVGVDRDHRPQKKAKGLGKEITLLMOTNLSTPPEKLAALCKKY 180
QY 197 AEILLEHRNSOKMILLQKQSQVQEKDHLRGHSHKAVLARSKLESLCRLQEHNRSLK 256
DB 181 AEILLEHRNSOKMILLQKQSQVQEKDHLRGHSHKAVLARSKLESLCRLQEHNRSLK 240
QY 257 EGVQARAREBEERKEKVTSHFOVTNDIQLQMEQHNRNSKLQENNELAERLKKLEQY 316
DB 241 EGVQARAREBEERKEKVTSHFOVTNDIQLQMEQHNRNSKLQENNELAERLKKLEQY 300
QY 317 ELREEHIDKVFHKDLQQLVDKLLQQAQEMLKEAEERHOREKDPLKRAVESORMCELM 376
DB 301 ELREEHIDKVFHKDLQQLVDKLLQQAQEMLKEAEERHOREKDPLKRAVESORMCELM 360
QY 377 KQETHLKKQALYTEKFEFQNTLSKSEVFTTFQEMEKMTKIKKLEKETTYRSRW 436
DB 361 KQETHLKKQALYTEKFEFQNTLSKSEVFTTFQEMEKMTKIKKLEKETTYRSRW 420
QY 437 ESSNKALLEMABEKTVRDKEGLQVQIKLQLEKLCRALQTERNDLNRVODLSAGGQSL 496
DB 421 ESSNKALLEMABEKTVRDKEGLQVQIKLQLEKLCRALQTERNDLNRVODLSAGGQSL 480
QY 497 TDSGPERRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQPTSARA 546
DB 481 TDSGPERRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQPTSARA 530

RESULT 7

US-09-976-740-8
; Sequence 8, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-8

Query Match 96.7%; Score 2693; DB 9; Length 530;
Best Local Similarity 99.8%; Pred. No. 3.1e-151;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAPRKPEGAQARTAQSGALRDVSEELS 76
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAPRKPEGAQARTAQSGALRDVSEELS 60

Qy 77 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 136
Db 61 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 120

Qy 137 KGDPTTEIRQSDVGRDHRHPQEKKAGLGKBITLLMOTLNTLSTPBEKLAALCKKY 196
Db 121 KGDPTTEIRQSDVGRDHRHPQEKKAGLGKBITLLMOTLNTLSTPBEKLAALCKKY 180

Qy 197 AELEEHRNSQKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNSLK 256
Db 181 AELEEHRNSQKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNSLK 240

Qy 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 316
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 300

Qy 317 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 376
Db 301 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360

Qy 377 KQETHLKKQALALYTEKFEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 436
Db 361 KQETHLKKQALALYTEKFEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 420

Qy 437 ESSNKALLEMAEKTVRDKELEGQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 496
Db 421 ESSNKALLEMAEKTVRDKELEGQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 480

Qy 497 TDSGPRRPEGPGQAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 546
Db 481 TDSGPRRPEGPGQAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 8

US-10-671-242-8
; Sequence 8, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; PRIOR FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-8

Query Match 96.7%; Score 2693; DB 12; Length 530;
Best Local Similarity 99.8%; Pred. No. 3.1e-151;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAPRKPEGAQARTAQSGALRDVSEELS 76
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAPRKPEGAQARTAQSGALRDVSEELS 60

Qy 77 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 136
Db 61 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 120

Qy 137 KGDPTTEIRQSDVGRDHRHPQEKKAGLGKBITLLMOTLNTLSTPBEKLAALCKKY 196
Db 121 KGDPTTEIRQSDVGRDHRHPQEKKAGLGKBITLLMOTLNTLSTPBEKLAALCKKY 180

Qy 197 AELEEHRNSQKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNSLK 256
Db 181 AELEEHRNSQKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNSLK 240

Qy 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 316
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 300

Qy 317 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 376
Db 301 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360

Qy 377 KQETHLKKQALALYTEKFEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 436
Db 361 KQETHLKKQALALYTEKFEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 420

Qy 437 ESSNKALLEMAEKTVRDKELEGQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 496
Db 421 ESSNKALLEMAEKTVRDKELEGQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 480

Qy 497 TDSGPRRPEGPGQAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 546
Db 481 TDSGPRRPEGPGQAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

497 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 546
481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 530

481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 530

RESULT 12
US-09-962-055-5
; Sequence 5, Application US/09962055
; Patent No. US20020052033A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,055
; FILING DATE: 24-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-962-055-5

RESULT 11
US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 96.7%; Score 2693; DB 15; Length 530;
Best Local Similarity 99.8%; Pred. No. 3.1e-151;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 17 KSSGPPGAGPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 76
Db 1 KSSGPPGAGPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 60
Qy 77 RQEDILSTYCVNNQGGPGEAGGPAEPEDAEKSTYVARGEPETPTVYNGEKEPS 136
Db 61 RQEDILSTYCVNNQGGPGEAGGPAEPEDAEKSTYVARGEPETPTVYNGEKEPS 120
Qy 137 KQDPNTEIROSDEVDGDRHRRPQKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 196
Db 121 KQDPNTEIROSDEVDGDRHRRPQKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Qy 197 AELLREHENSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLARSKLESICRSLQRHNSLK 256
Db 181 AELLREHENSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLARSKLESICRSLQRHNSLK 240
Qy 257 EGVQARAEERKREKVTSHFQVTLNDIQLOMEQHNSKLRQNMELAEKLLIROY 316
Db 241 EGVQARAEERKREKVTSHFQVTLNDIQLOMEQHNSKLRQNMELAEKLLIROY 300
Qy 317 ELREBEHIDKVPKHDLQOOLVDKALQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
Db 301 ELREBEHIDKVPKHDLQOOLVDKALQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Qy 377 KQOETHLKOALYTEKPEFQNTLSKSEVPTTFKQEMKRTKKIKLEKETMYRSRW 436
Db 361 KQOETHLKOALYTEKPEFQNTLSKSEVPTTFKQEMKRTKKIKLEKETMYRSRW 420
Qy 437 ESSNKAILEWAEKTVRDELSGLQVKIQRLKLCALQTERNDLNKRQDLSAGQGSLL 496
Db 421 ESSNKAILEWAEKTVRDELSGLQVKIQRLKLCALQTERNDLNKRQDLSAGQGSLL 480

Query Match 91.7%; Score 2552.5; DB 9; Length 557;
Best Local Similarity 91.0%; Pred. No. 6.7e-143;
Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;
Qy 1 MNQODKNGAKQSNPKSSPCQPEAGPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 60
Db 1 MNQODKNGAKQSNPKSSPCQPEAGPEGGAQAPSSPRVTEAPCYPGAPSTASQOTGPQEPSTARA 59
Qy 61 RTAQSGALRDVSEHLSRQLEDILSTYCVNNQGGPGEAGGPAEPEDAEKSTYVARN 120
Db 60 RTAQSGALRDVSEHLSRQLEDILSTYCVNNQGGPGEAGGPAEPEDAEKSTYVARN 119
Qy 121 GEPEBP-TPVNGEKEPKGDPNTEIROSDEVDGDRHRRPQKKKAKGLGKEITLLMQTL 179
Db 120 GEPEBP-TPVNGEKEPKGDPNTEIROSDEVDGDRHRRPQKKKAKGLGKEITLLMQTL 179
Qy 180 NTLSTPEKLAALCKKYAELEERHNSOKMKLLQKQSQOLVQEKDHLRGEHSKAVLARS 239

Db 399 DTATWKSFRFNCNRRALLDMIEKAMRTKEVCFVLKIQRLNLCRALQBERNELYRKIK 457

RESULT 2
F87844

Protein T22C1.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: F87844
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F87844
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:chr_1; PIDN:CAA99923.1; PID:G93880029; GSPDB:GN00019; CESP:T22C1.6
C:Genetics:
A:Gene: T22C1.6
A:Map position: 1

Query Match 13.5%; Score 375; DB 2; Length 346;
Best Local Similarity 30.8%; Pred. No. 5e-09;
Matches 105; Conservative 69; Mismatches 123; Indels 44; Gaps 9;

QY 166 KGLGKEITLIMQ-----TLNLT-STPE-EKLAALCKKYAELEHNRNSQOMKLLQKQ 217
Db 3 KNFKLISPIFNEGDEAALLKSLGVPDAEKVQKLIKLA--SEKQNAELKIKVLDYDK 60

QY 218 SOLVQKHRLREHSAVL-----ARSKLESICRLQHRNRLKEEGVQVAREBEKKE 272
Db 61 VVKVLDLFEKLRNQIILLTTEAKSLKLEELCRGLQKANHOTREACAKMKLEVERGL 120

QY 273 VTSHPQVTLNDIQLOMEHNRNSKLRQENMELARLKLIOYB-----LRBE 321
Db 121 AVSQLVKILDKLEKTWAGRSKSDSLAEDNKKLSKFSFGHYEKMKVDDQIQKEK 180

QY 322 HDVVKFKHDLQQLVDAKLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOET 381
Db 181 YWEEYGTNDLILKLTAKLESASIQVKS-----GMEKDELAKIMLEETARVGGALKTEK 236

QY 382 HLKQQLALYTERFEFFQNTLSKSSVFTTFKQEMKMTKIKKLEKTTMYRSWESSNK 441
Db 237 ALREQVQVSAKYSILTSCLSKSNEAFDKFDEISRNVNKKMQVKEGLSYKKSDEANK 296

QY 442 ALL-----EKAEKTVRDKLEGLQVKIQRLKLCRALQ 475
Db 297 KVLVLTMTNOEYAEKIATSDK-----KIQMLENLCLALR 330

RESULT 3
T25110

Hypothetical protein T22C1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25110
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19982
A:Accession: T25110
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-335 <WIL>
A:Cross-references: EMBL:Z75550; PIDN:CAA99923.2; GSPDB:GN00019; CESP:T22C1.6
A:Experimental source: clone T22C1
C:Genetics:
A:Gene: CESP:T22C1.6
A:Map position: 1
A:Introns: 6/1; 46/3; 66/3; 91/3; 285/2

Query Match 13.4%; Score 372.5; DB 2; Length 335;
Best Local Similarity 30.8%; Pred. No. 6.1e-09;
Matches 103; Conservative 70; Mismatches 120; Indels 41; Gaps 9;

QY 168 LGKEI--TLLMQTLTSTPE-EKLAALCKKYAELEHNRNSQOMKLLQKQSQOLVQEK 224
Db 1 MGKFNDEAALLKSL--GVPDAEKVQKLIKLA--SEKQNAELKIKVLDYDKVVKVQDL 56

QY 225 DHLRGESKAVL-----ARSKLESICRLQHRNRLKEEGVQVAREBEKKEKVTSHQV 279
Db 57 TEKLERNNQTLTTEAKSLKLEELCRGLQKANHOTREACAKMKLEVERGLAVEQLKV 116

QY 280 TLNDIQLOMEHNRNSKLRQENMELARLKLIOYE-----LPEEHIDKVPK 328
Db 117 TLXDIKTWAGRSKSDSLAEDNKKLSKFSFGHYEKMKVDDQIQKEKKEKWEYVK 176

QY 329 HKDLQQLVDAKLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKQOLA 388
Db 177 TKOLEIKLLTAKLESASIQVKS-----GMEKDELAKIMLEETARVGGALKTEKALREQVQ 232

QY 389 LYTEKFEFPNTLSKSSVFTTFKQEMKMTKIKKLEKTTMYRSWESSNKALL-----444
Db 233 EYSAKYSILTSCLSKSNEAFDKFDEISRNVNKKMQVKEGLSYKKSDEANKKVLVITM 292

QY 445 ---EMAEKTVRDKLEGLQVKIQRLKLCRALQ 475
Db 293 TNQYAEKIATSDK-----KIQMLENLCLALR 319

RESULT 4
T46337

Hypothetical protein DKF2p43402413.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46337
R:Kocher, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23037
A:Accession: T46337
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-992 <AAA>
A:Cross-references: EMBL:AL137265
A:Experimental source: adult testis; clone DKF2p43402413
C:Genetics:
A:Note: DKF2p43402413.1

Query Match 10.1%; Score 281; DB 2; Length 992;
Best Local Similarity 22.1%; Pred. No. 0.00011;
Matches 138; Conservative 103; Mismatches 213; Indels 170; Gaps 21;

QY 12 KQNPXSSPG-----QPEAGPEGAGQERPSQAAPVAREGP-----GSSQAPKPECAQA 60
Db 87 KEENEKSEPKICRNLTVPKADPTG-----SEPAKASEKEAPETVDAGEGSRREAAKE 141

QY 61 RTAQSGALRDVSBELSRQLE-----D 81
Db 142 PKKASALEEGSSDASQELISEHMKPEQLSDSIASDPKSFHGLDFGFRSRISEHLLDVD 201

QY 82 ILSTY-----CVDANQGGPDGAGQGPAPPE-DAEKSRITYVANGEP-----EPTPV 128
Db 202 VLSFVLGACRQAQQPLGTHDDKDDSSQDELQSKGLERLSPLPHEERAQSPRS 261

QY 129 VNGEKPSKGDPNTEIROSDRVDG-----153
Db 262 LATEEPFQGPQEPQEPWEKAEELGSDSASLSLQSLQREQAPSPFAACEKGEHQSOAE 321

QY 154 -----RDHRRPQKKVAKGLGKEITLIMQTLNLTSTPE-----KLAALCKKYAELEHNR 204
Db 322 ELGFGQSEADPEEKVAVSPPTFPVSPEVSTSTPEVAPPEQLSEALKAMEEAVQVLE---378

QY 205 NSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLA--RSKLESICRLQHRNRLKEGVOR 262

```

Db 379 --QDORHLESQKQMOOLREKLQOESEEBITLRLHQKQESLSSLRERLQKAEIEBA-R 435
Qy 263 ARBEEK-----RKEVTSHFQVTLNDIQLOME-----QHNRNSKLRQENNEL 305
Db 436 MREESQRLSWLRAQVQSSTQADEQIRAEQEAASLQKLREBESQQAERASLEQKQRQM 495
Qy 306 AERLKKLIEQVELBEHIDKVPKHDLQO-----OLVDAKLOQAQOEMLXE-----A 351
Db 496 LEQKUEBIEASEQEAALNAKAKALQOUREQLEGERKEKAVATLEKESAEERLCSL 555
Qy 352 EERHQRKDFLLKEAVESQRCMLMKO-----QETHLQKQALALYTEKEFEFQNTL--SKSS 405
Db 556 EAKHREVVSLLQKIQEAQKEBAQKLOKCLQGVHVRHVSQYHVGVEHLSLLREKQ 615
Qy 406 EVFTTFQEMKTKKIKLEKETMYRSRWESSNKALLEMAEKTVRDKELEQLQVKIQ 465
Db 616 EVEGEHERLLDKM-----KEHQVMAKAREQYE-----AEERKQASILLGLHGLE 663
Qy 466 RLEKLC-RALQTERNDLNKRVODL 488
Db 564 RQRAHERLETVRQEQKLEDL 687

RESULT 5
T50985
related to transcription factor TMP [imported] - Neurospora crassa
N:Alternate names: protein B7F18.40
C:Species: Neurospora crassa
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: T50985
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-892 <SCH>
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.40
A:Experimental source: BAC clone B7F18; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F18.40
A:Map position: 6
A:Introns: 63/1; 551/3

Query Match 10.0%; Score 278; DB 2; Length 892;
Best Local Similarity 23.2%; Pred. No. 0.00013;
Matches 142; Conservative 106; Mismatches 247; Indels 118; Gaps 27;

Qy 2 KQDKKNGAAQSNPKSPGQPEAGPEGACERPS---QAPAVEAGPGSSQAPRKEGA 58
Db 91 KNL-D-RTSSSTQASPRQSMAPSRASSTSIERNFADKASFNNAVSPRASADTPKRSQDT 149
Qy 59 --QARTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQGPAPDEAKSRTY 116
Db 150 TQEPVPAPLDSADDIKEVAQSVSE-----XADSRPSTD-QPSPENTEDEAPT--- 197
Qy 117 VARGNEPPTPVVNGEKEPSKGDNTREIROSDVGDGRDHRROEKKKAGLKEITLIM 176
Db 198 --TESKLEAPAEILAKTEPKSEVETKARPQDQEMKSDDRPCNQDEIYAYVERIDALE 255
Qy 177 QTL-----NTLSTP-----BEKLAALCKKVAELLEERN--SOKO-----MKL 212
Db 256 AKLQYLAREASAAARKKALSPAGSAEKKLAKDQQLAQIMEKKNLASNEQKLTILKN 315
Qy 213 LQKQSQOLVQEKDHLRGEHGAVKLARSKLESCLR--ELQRHNSRLKEGVQVAREEEK 269
Db 316 LRKQQAEDKONGNLKAAKAKADREIENLRKRARHAELEKSSQVQLKRLDQSQRELYNL 375
Qy 270 RKEVTSHFQVTLNDIQLOMEQHNER-----NSKLRQENMELABELKKLIEQVELRE 321
Db 376 RPEVSK-DTIIARLSQIQKATQADVMASAKANDKAREQQQ-----RRIAL-----EE 424
Qy 322 HIDKVPKHDLQOQLVDKLOQAQOEMLKAEERHQREKDF-----LLKEAVESQR- 371

```

```

Db 425 SVEALKIEKNL---MADRAQAQADELRKEAKERAKASEKAKALELKAEBVHNMESKLEAMRT 481
Qy 372 -----MCELMKQOETHLQKQALALYTEKEFEFQNTLSKSSEVFTTFKQEMK 418
Db 482 RAEAEASSGYTQDSQAKLLAQVET-LQSQYSIASENWQGIETTL--RSRIVNLSKERDE-- 536
Qy 419 TKKIKKLEKETMYRSRWESSNKALLEMAEKTVRDKELEGLQVKIQRLKELCR-----AL 474
Db 537 -----ALQRESNRRKAREAE---LEBAKTKLPNOEDVESYRSQDLSLKKRAEAEAL 588
Qy 475 QTERNDLNKRVODLSAGGOSLTDSPERRPEGGAQAPSSPRVTEAPCYPGAP--STEA 532
Db 589 AERADPEKQKQAWEA--EKELIKEEREDLQSQG-----NRER-SWLEGLPGGPFLLKNEG 641
Qy 533 SGQTGPQRPPTSAR 545
Db 642 SGPGSPQLSTAQR 654

RESULT 6
A38194
desmoplakin I - human
N:Contains: desmoplakin II
C:Species: Homo sapiens (man)
C>Date: 04-Mar-1993 #sequence_revision 03-May-1996 #text_change 21-Jul-2000
C:Accession: A38194; A35536; B35536
R:Virata, M.L.; Wagner, R.M.; Parry, D.A.; Green, K.J.
Proc. Natl. Acad. Sci. U.S.A. 89, 544-548, 1992
A:Title: Molecular structure of the human desmoplakin I and II amino terminus.
A:Reference number: A38194; MUID:92115697; PMID:1731325
A:Accession: A38194
A:Molecule type: mRNA
A:Residues: 1-974 <VIR>
A>Note: sequence extracted from NCBI backbone (NCBIN:75984, NCBI:P:75986)
R:Green, K.J.; Parry, D.A.D.; Steinert, P.M.; Virata, M.L.A.; Wagner, R.M.; Angst, B.D.
J. Biol. Chem. 265, 2603-2612, 1990
A:Title: Structure of the human desmoplakins. Implications for function in the desmosome
A:Reference number: A35536; MUID:90153880; PMID:1689290
A:Accession: A35536
A:Molecule type: mRNA
A:Residues: 'R', 927-2677 <GRE>
A:Cross-references: GB:J05211; NID:gl81607; PIDN:AAA35766.1; PID:gl81608
A:Accession: B35536
A:Molecule type: mRNA
A:Residues: 'R', 927-1000, 1600-2677 <GR2>
A:Cross-references: GB:J05211
C:Comment: Desmoplakins I and II, products of single gene, are major proteins of the in
C:Genetics:
A:Gene: GDB:DSP
A:Cross-references: GDB:126564; OMIM:125647
A:Map position: 6pter-6p21
C:Keywords: alternative splicing; cytoskeleton
F:1-2677/Product: desmoplakin I #status predicted <MAT1>
F:1-1000, 1600-2677/Product: desmoplakin II #status predicted <MAT2>

Query Match 9.4%; Score 263; DB 2; Length 2677;
Best Local Similarity 23.4%; Pred. No. 0.0016;
Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 25;

Qy 42 EAEFGSSQAPRKP-----EGQAQRTAQSS-----GALRDVSELSRQLEDIL 83
Db 1238 QKATGSEVSRQKQLEVLQVQVQRTSESVRYKQSLDAAKTIQDKNKEIER-LKQLI 1296
Qy 84 STYCVNNQGGPGEDGAQGPAPDEAKSRRT-YVARGNEPPTPVN-----GEKEPSKG 138
Db 1297 DKETNDR-----KCLEDEVARLQRVQYDLQKANSATETINKLRVQEQLTEL 1344
Qy 139 DPNTEBIRQSDVGDGRDHRROEKKKAGLKEITLIMQTLSTLSTPEEKLAALCKKYAR 198
Db 1345 RIDYEVRSERTVKQDQITRFQNSLKELOLQKQ--KVREELNRLKXTASEDCSKKLE 1402
Qy 199 LLEEHRNSQKMKL-----LQKQSQ--LVQEKD-----HLR-GHSHKAVLAR 238

```

Db 1403 ELEGHRSLSKEQAIKTNLTQLEQASIVKKSSEDLQQRDVLGDHLEKQRTQBELR 1462
Qy 239 --SKLESCLRELORHNSLKE--EGVQAPBEEBKEKVTSHFQVTLNDIQLOVEQH 291
Db 1463 LSSEVEALRQLQESVKAHLNHPKAIEDKSR-----SLNESKIBERL 1512
Qy 292 NERNSKLQENNELAERLKL--IEQYELR-----BEHIDK-----VFKHQLOQQ 334
Db 1513 QSLTENLTKEHLMLEELNLRLEVDLRRGRSEADSKNATILELSQLQISNNRTLEL 1572
Qy 335 QLVDAKLOQAOBMLKEAERHQREKDFLLKRAVESORMC-ELMQQETHLKKQALITYEK 393
Db 1573 QGLINDLQRENLQOEIKTQKQALEASNIQESKQCTQVQBERSLLVKIKVLEQDK 1632
Qy 394 -----FEFQNTLSKSSSVFTTPEKQEMKTKIKK-LEKETTMYRSWSSSNKALLEMA 447
Db 1633 ARLOLEDELNRAKSTLEATFTVKORLECEKQIQNDLNQWTKYYSRKEBAIRK--IEST 1690
Qy 448 BEKTVRDK-----ELSGLOVKIQRIEKLOR-----ALQTERNDLNKRVQDLISA 490
Db 1691 REKSPREKNSLSETERLQAEIKRIEERCRAKLEDSTRETQSQLETSRYQREIDKL-- 1748
Qy 491 GGQGSLSLDSGPERPEGPGCAQAPSSPRVTEAPC 523
Db 1749 -----RQRPYG-----SHRETQTEC 1763

RESULT 7
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
R:Accession: S28589
R:Fietz, M.J.; Rogers, G.B.
A:Submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <PIE>
A:CROSS-references: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
P:49-81/Domain: calmodulin repeat homology <EP2>

Query Match 9.4%; Score 261.5; DB 1; Length 1407;
Best local Similarity 23.5%; Pred. No. 0.00097;
Matches 117; Conservative 110; Mismatches 168; Indels 103; Gaps 22;

Qy 42 EAGPGSS-----QAPKPGGAQRTAQSGALRDVSEELSRQLEDILSTYCVDNQGG 94
Db 514 QEQRPQGTWRWQLQBEAQRRHTLYAKPGQQLRE-BEELQ-----BKRRQE 561
Qy 95 PGEDGAQGPAPDEPAKSRVTYVARGPEPTPVNGEKPSPKDPNTEERQSDVEGDR 154
Db 562 RERETVEELQREBEKRR-----RQERQYRLEELQBEQLADR 604
Qy 155 DERRPQEKKAQGLGKHTLMTLNTLTSTPEKLAALCKKYAEALLEE-HRNSQKMKLL 213
Db 605 KLREBEQLQOER---EERLRQERERKLREBEQL--LRQEQELRQERERKLREBEQL 659
Qy 214 QKKQQLVQEKDH-LRGEHSKAVLARSKLCSLREBLQHRNLSLKEGVQVAREES--EK 269
Db 660 RREBEQLRQERERKUREEQ-----LIQREEBELRQERARKLREBEQLRQER 716
Qy 270 RKEVTSHFQVTLNDIQLOMEQHNRNSKLQENNEL-----AERLKLIEQYELREHIDK 325
Db 717 ERKLREBEQLRREBEQLLQ---ERDRKLREBEQLQESSEBRLRQERQQLRRER-DR 772

Qy 326 VFHKD-----LQOQLVDKLOQAOEMLKEAE-----ERHQEKDFLLKEA 366
Db 773 KFBEEQLQOERERERKLREBEQLQOERERERKLREBEQLRQERERKLREBEQLQOER 832
Qy 367 VESQRC-----ELMQQETHLKKQALITYEKPEFQNTLSKSSSVFTTPEKOE 414
Db 833 -EERLRQERERKLREBEQLRQEQELRQESA---RKLREBEQLRQEQ---ELRQE 885
Qy 415 MEKWTQKIKKL---EKETTMYRSWSSSNKALLEMAEKTIVRDKEGLGVQIKQRIEKL 470
Db 886 RDRKLREBEQLRQEQELRQERDRKLREBEQLQOERERERKLREBEQLRQERER 942
Qy 471 CRALQTS-RNDLNKRVQD 487
Db 943 LRREBEQLRERARKLRE 960

RESULT 8
H69378
conserved hypothetical protein AF1032 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: H69378
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.B.; Ketchum, K.A.; Dods,
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69378
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-886 <KLE>
A:CROSS-references: GB:AE001032; GB:AE000782; NID:g2689355; PIDN:AAB90211.1; PID:g2649;
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 9.3%; Score 260; DB 2; Length 886;
Best local Similarity 24.6%; Pred. No. 0.0007;
Matches 118; Conservative 90; Mismatches 118; Indels 154; Gaps 23;

Qy 108 EDAEKSTYVARGPEPTPVNGEKPSPKDPNTEERQSDVEGDRHRRPOEKQKAG 167
Db 155 ENAWKNLGAVIR-----MLEREKERLK-----EFLSQEEQI---KRQEKKA-- 194
Qy 168 LGKEITLLMTLNTLTSTPEKLAALCKKYAEALLEEHRNSQKMKLLQKQSLVQEKDHL 227
Db 195 ---EIERISBEIKSIESLREKLS-----EVRNLSRLKELEHEKLSRL-----ESL 237
Qy 228 RGEHSKAVLARSKLCSLREBLQHRNLSLKEGVQVAREEBEKEKV----- 273
Db 238 RKQESSVLQVGRGLEEKLELE---KQKEV-VERIEDLEKAKEVKELPKABRYSILE 293
Qy 274 --TSHFOVTLND-----IQLOMEQHNRNSKLQENNELAERLKLIEQY 317
Db 294 KLASEINQALRDVEKRGDLTREAAAGTQALQKAEHNSKL-EETIKRIELELERERPE 352
Qy 318 -----LREBHI---DKVPKHKDLQOQLVDKLOQAOEMLKEAE 353
Db 353 KSHRLLETLPKMDRMQGIKAKLEENLTPDKVEKYDL-----LSKAKEEKEITE 404
Qy 354 RHQR---BKDFL-----LKEAVE---SORMELAKQO---ETHLKKQALITYEKPEFQ 398
Db 405 KLKGLIAKGSLSLTRGAQLKKAVEELKSAERTCPVCGRELDDEHRRKNIMABYTEMRIA 464
Qy 399 NTLKSSESVFTTPEKQEMKTKIKKLEKETTMYRSW-----ESSNKALL 444
Db 465 BELAQDAIEKKLERKVE---KALEKETVLKRWDELKALENELSSHDAEKLSA 521
Qy 445 EMABEKTVRDKLEGL--QVKI-----ORLEKLCRALQTERNDLNKRVQD 487

A;Residues: 1-205 <HAY>
 R;Komine, Y.; Maita, T.; Matsuda, G.
 J. Biochem. 110, 60-67, 1991
 A;Title: The primary structure of skeletal muscle myosin heavy chain: II. Sequence of the
 A;Reference number: PX0051; MUID:92041768; PMID:1939028
 A;Accession: PX0051
 A;Molecule type: protein
 A;Residues: 206-636 <KOM>
 R;Maita, T.; Miyaniishi, T.; Matsuzono, K.; Tanioka, Y.; Matsuda, G.
 J. Biochem. 110, 68-74, 1991
 A;Title: The primary structure of skeletal muscle myosin heavy chain: III. Sequence of the
 A;Reference number: PX0052; MUID:92041769; PMID:1939029
 A;Accession: PX0052
 A;Molecule type: protein
 A;Residues: 201-213;632-837 <MAI>
 R;Maita, T.; Yajima, E.; Nagata, S.; Miyaniishi, T.; Nakayama, S.; Matsuda, G.
 J. Biochem. 110, 75-87, 1991
 A;Title: The primary structure of skeletal muscle myosin heavy chain: IV. Sequence of the
 A;Reference number: JX0178; MUID:92041770; PMID:1939030
 A;Accession: JX0178
 A;Molecule type: protein
 A;Residues: 833-1938 <WA2>
 R;Maita, T.; Hayashida, M.; Tanioka, Y.; Komine, Y.; Matsuda, G.
 Proc. Natl. Acad. Sci. U.S.A. 84, 416-420, 1987
 A;Title: The primary structure of the myosin head.
 A;Reference number: A26365; MUID:87092420; PMID:3467365
 A;Accession: A26365
 A;Molecule type: protein
 A;Residues: 1-129, 'X', 131-139, 141-550, 'X', 552-754, 'X', 756-784, 'QL', 787-804, 806-810 <MA3>
 R;Watanabe, B.
 Biol. Chem. Hoppe-Seyler 370, 55-61, 1989
 A;Title: Amino-acid sequence of the hinge region in chicken myosin subfragment-2.
 A;Reference number: S02082; MUID:89228549; PMID:2713098
 A;Accession: S02082
 A;Molecule type: protein
 A;Residues: 1144-1270 <WAT>
 R;Yajima, E.
 Nagasaki Igakkyo Zasshi 65, 409-430, 1990
 A;Title: Study on tail region of skeletal muscle myosin; primary structure and protease
 A;Reference number: PW0009
 A;Accession: PW0009
 A;Molecule type: protein
 A;Residues: 1304-1938 <YAJ>
 R;Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
 submitted to the EMBL Data Library, August 1991
 A;Description: Analysis of the chicken fast myosin heavy chain family: Localization of
 A;Reference number: S39081
 A;Accession: S39081
 A;Molecule type: mRNA
 A;Residues: 1081-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1795, 'HV', 1798-1829, 'S', 1831
 A;Cross-references: EMBL:M74084
 R;Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
 J. Mol. Biol. 225, 1143-1151, 1992
 A;Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
 A;Reference number: S24348; MUID:92309413; PMID:1377278
 A;Accession: S24351
 A;Molecule type: mRNA
 A;Residues: 1082-1182, 'T', 1184-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1780, 'M', 1782-
 A;Cross-references: EMBL:M74084
 R;Watanabe, B.
 Biol. Chem. Hoppe-Seyler 370, 1027-1034, 1989
 A;Title: Complete amino-acid sequence of subfragment-2 in adult chicken skeletal muscle
 A;Reference number: S05515; MUID:90121764; PMID:2610940
 A;Accession: S05515
 A;Molecule type: protein
 A;Residues: 842-906, 'Q', 908-1270 <WA3>
 R;Watanabe, B.
 Biol. Chem. Hoppe-Seyler 370, 549-558, 1989
 A;Title: Amino-acid sequence of the short subfragment-2 in adult chicken skeletal muscle
 A;Reference number: S04501; MUID:89374803; PMID:2775482
 A;Accession: S04501
 A;Molecule type: protein
 A;Residues: 852-906, 'Q', 908-1108 <WA2>

R;Matsuda, G.; Maita, T.; Miyaniishi, T.; Hayashida, M.
 J. Protein Chem. 6, 33-46, 1987
 A;Title: Structure and function of muscle myosin.
 A;Reference number: A60877
 A;Accession: A60877
 A;Molecule type: protein
 A;Residues: 1-139;141-205 <MA1>
 R;Gulick, J.; Kropp, K.; Robbins, J.
 J. Biol. Chem. 260, 14513-14520, 1985
 A;Title: The structure of two fast-white myosin heavy chain promoters. A comparative at
 A;Reference number: A32507; MUID:86033956; PMID:2997212
 A;Accession: A32507
 A;Molecule type: DNA
 A;Residues: 'X', 1-168 <GUL>
 A;Cross-references: GB:M13512; GB:M12083; GB:M13510; NID:G212363; PIDN:AAA49866.1; PID:
 R;Kropp, K.; Gulick, J.; Robbins, J.
 J. Biol. Chem. 261, 6613-6618, 1986
 A;Title: A canonical sequence organization at the 5'-end of the myosin heavy chain gene
 A;Reference number: A32587; MUID:86196091; PMID:3009465
 A;Accession: C32517
 A;Molecule type: DNA
 A;Residues: 'M', 1-56, 'T', 58-76, 'I', 78-168 <KRO>
 A;Cross-references: GB:M13515; GB:M13511; NID:G212373; PIDN:AAA48971.1; PID:9555468
 C;Comment: This is a fragment of the globular head.
 C;Superfamily: myosin heavy chain; myosin motor domain homology
 C;Keywords: acetylated amino end; actin binding; ATP; coiled coil; hydrolase; methyle
 P;1-1938/Product: myosin heavy chain #status experimental <MAT>
 P;1-1938/Domains: myosin motor domain homology <MYOT>
 P;89-768/Domains: myosin motor domain homology <MYOT>
 P;179-186/Region: nucleotide-binding motif A (P-loop)
 P;550-587/Region: actin binding #status predicted
 P;657-679/Region: actin binding #status predicted
 P;841-1938/Domains: coiled coil <COI>
 P;841-1289/Region: S2
 P;852-1108/Domains: short subfragment 2 <SUB2>
 P;1290-1938/Region: light meromyosin
 P;1/Modified site: acetylated amino end (Ala) #status experimental
 P;130,551/Modified site: N6-methyllysine (Lys) #status experimental
 P;185/Binding site: ATP (Lys) #status predicted
 P;697,707/Active site: Cys #status predicted
 P;755/Modified site: 3'-methylhistidine (His) #status experimental

Query Match 9.1%; Score 254; DB 1; Length 1938;
 Best Local Similarity 22.3%; Pred. No. 0.0037;
 Matches 122; Conservative 83; Mismatches 169; Indels 174; Gaps 19;
 QY 49 SQAPKPEGAQARTAQSGALRDVSELSRQLEDILSTYVDNNQCGPGDGAQGPAPPE 108
 DB 1090 SOIQSKIHEQDLGQ---LQKKIKELQARTIEL---EEIEAERTSA 1132
 QY 109 DAESRTYVARNGEPEPTPVVNGEKEPSKGDPTTEIRQS-DEVG-----DRDRRPQ 160
 DB 1133 KAEKRAILSR-----ELSEISERLEFAGGATAAQTENMKREA 1171
 QY 161 EKKKAGLGGKETLLMQTLNTLSTPEKLAALCKV---AEILLSEHNSQOMKLLQKK 216
 DB 1172 EFQKWRDLDEATL-----QHEATAALRKHAOSTAELGSCIDMLQVKQKLEKE 1222
 QY 217 QSQLVQEKDHLRGHSHKAVLARSKLESLEL-----QRHRSLSKEGVQAR- 264
 DB 1223 KSELNMEIDDLASNMESVSKAKANLEKMCRTLEDQLSEIKTKEQNMINDLNTQARL 1282
 QY 265 -----EHEEKKEVTSH-FOYTLNDI 284
 DB 1283 QTETGEYSRQAEKDALISQLSRGQGTQQTTEELKRLHEEIRIKAKNALAHALQASRHC 1342
 QY 285 QLQMEQHNER-----NSKLR-----QENMELAEKLLKLEQVELR 319
 DB 1343 DLLRQYEEQAKGELQALSKANSEVAQWTKVETDAIQTEELBAKKLQALQIDA 1402
 QY 320 BEHIDKVFHKDLQOOLVDAKLQQQAEMLKEAEERHQREKDFLKEAVESQMCHELMKQ 379
 DB 1403 BEHV-----EAVNAKCA-----LEKTKORLQNEVEDLVYDVERSNAAALDKK 1447


```
QY 380 ETHLKQQLALYTEKEEFBNQTL-----SKSSSEVF---TFEQK---EMEKTKKIKKL 425
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1448 QNFKILIAWKQKTEETQTELEASQKESRUSLSTELFKKKNAYEBSLDHLSTLAKENKL 1507
QY 426 EKETTMYRSWESSKALLEMAEKTVDKLEGLQVKIQBLE-----KLCLALQTE 477
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1508 QOETADLTQIAEGGKAVHELEKVKKHVEQEKSELQAALEBAEASLEHEGKILR-LQLE 1566
QY 478 RNDLNKRV 485
D 1567 LNQIKSEI 1574

RESULT 14
A56734
ribosome receptor, 180k - dog
C:Species: Canis lupus familiaris (dog)
C>Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 08-Oct-1999
C:Accession: A56734
R:Wanker, E.E.; Sun, Y.; Savitz, A.J.; Meyer, D.I.
J. Cell Biol. 130, 29-39, 1995
A>Title: Functional characterization of the 180-kD ribosome receptor in vivo.
A:Reference number: A56734; MUID:95310363; PMID:7790375
A:Accession: A56734
A>Status: Preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1534 <WAN>
A:Cross-references: GB:X87224; NID:g984113; PIDN:CAA60676.1; PID:g984114
C:Keywords: endoplasmic reticulum; membrane protein; protein biosynthesis
F:198-743/Region: 10-residue repeats (N-Q-G-K-A-E-G-A-P)

Query Match 9.1%; Score 253.5; DB 2; Length 1534;
Best Local Similarity 20.7%; Pred. No. 0.0023;
Matches 158; Conservative 115; Mismatches 229; Indels 261; Gaps 31;

QY 3 NQDKKNGAAKQSNPKSSPGQEPAGE-----GAQRRPQAA 38
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 697 NQKK-GBCTNQGKSSGNOGKVDASANQSKRASPLOGKNADWVQSQAPKQEA 755
QY 39 PAVRABGPGSSQAPKP-----EGAQR-----TAQSGALRD 70
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 756 PAKKSGSKGEGEPGPPSDSPLPYKTLVSTVGSVMVFNEGEARLJIELSEKAGVIQD 815
QY 71 VSEB-----LSROLED-----ILSTYCVNNQGGPGDGAQGEPAEPDEAKSRT-- 115
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 816 THKATQKGDPAVKLRQLEKEKLLAT-----EQDAVAASKL 855
QY 116 -YVARGPEPTPVVNGEKPSK-----GDPNTEIRQSDVGDHRDPQEKKAK---- 166
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 856 RVNKLAAEAKAAAGKAKVKQKVAREQBITAVQARIEASYREHVKEVQQLQCKIRTL 915
QY 167 -----GLGKEITILMQTLN-----TLSTPEEKLAALCKKVAELLE 201
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 916 QEQLNGNTQLARLQGENSILRDALNQTQSVSKQNTLAKLRQELSKVSKSELVERSE 975
QY 202 EHRNSQKMLQKKQKQSVQEKDHLRGEHSKAVLA-RSKLESICREL---QRNRLSKE 257
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 976 AARQEEQKKALETKTAALEKQVLQASHKSEBEALQKRLDEVSRELCSLTSHASLRA 1035
QY 258 EGVQBARBEERKREVTSHFQVTLNDIQLOMEQHNRSKL---RQENMELARLKK--- 311
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1036 D-AEKAQEQQQAALHLSKLSQSSRAEVKSKSEELSGHLQKEAPANSQLMIRISIEA 1094
QY 312 LIEQVELAE-----EHTDKVFKKDLQQLQVDAKLQQAQKMLKEABERHO----- 356
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1095 LLEAGQARDTQDAQASRAEHQARL---KELSSQV---WCLEKEATELKEAVEQCKVKNDL 1149
QY 357 REKDFLLKEAVES-QRMCLMKQOETHLKQ-----LAL----- 389
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1150 REKNWKAMEALASARACEEKLRLSUTQAKESQKLSLITEAQTKKALLALLPALSSAPQ 1209
QY 390 -YTE-----KFEFQNTLSKSESVFTTFQEMEK 417
```

```
D 1210 SYTEMQLQELREKPELLKQRPADTDSSDLASKLREAEFTQNNLQAECDQYRILATEG 1269
QY 418 MTKKIKK-LEKETTMYRSWESSKALLEMAEKTVDKLEGLQVKIQ----- 465
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1270 MLKDLQKSVSEEBQV---WKAIVSATSEELQKSRVTVKHLEDIVKLGKLELSEBQVRE 1325
QY 466 -----RLKELCRALQTERNDLNKRVQDLSAGGQSLTDSGPPRRPEGPAQAAPSS--- 515
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1326 HTSHLEALEKHWAAASAEQSYAKEV-----AGLRQLLESQSLDAAKSEAQKSNELA 1381
QY 516 ---PRVTEAPCY-----PGAPSTASGQTGPQEPPTSARA 546
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1382 LVRQLSEMKSHVEDGDVAGSPAAPPAAE-----QDPVELKA 1417

RESULT 15
A40691
trichohyalin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A40691; A34209; S32633
R:Pietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A>Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding
A:Reference number: A40691; MUID:93260018; PMID:7684041
A:Accession: A40691
A:Molecule type: DNA
A:Residues: 1-1549 <PIE>
A:Cross-references: EMBL:Z18361; NID:g295940; PIDN:CAA79165.1; PID:g295941
A>Note: sequence extracted from NCBI backbone (NCBI:P:132511)
R:Pietz, M.J.; Presland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A>Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker
A:Reference number: A34209; MUID:90130632; PMID:2298812
A:Accession: A34209
A:Molecule type: mRNA
A:Residues: 1016-1151, 1205-1257, 1281-1398, 'G', 1400-1549 <PI2>
A:Cross-references: GB:X51695; NID:gi827; PIDN:CA335992.1; PID:gi828
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
A>Note: single copy gene
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <BF2>
F:387-851/Region: 28-residue repeats
F:886-1519/Region: 23-residue repeats

Query Match 9.1%; Score 253.5; DB 1; Length 1549;
Best Local Similarity 22.7%; Pred. No. 0.0023;
Matches 111; Conservative 108; Mismatches 194; Indels 77; Gaps 17;

QY 22 QPAGPEGQRRPSQAAPAVEAGPGSSQAAPKPEGAQARTAGSGALRDVSELSQLED 81
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 521 QREERKRRERQRYLEKVELQEEBQLQREKREKQREKQ-----YLEKVELQEE 574
QY 82 ILSTYCVNNQGGPGDGAQGEPAEPDEAKSRTYVARGPEPTPVVNGEKPSKGDPN 141
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 575 QL-----QREKRRERQRYLEKVELQEEBQLQREKREKQRE 617
QY 142 TE-----EIRQSDVGDHRDPQEKKAGLGEKTEILLMQTLNTLSTPEEKLAALCKKY 136
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 618 RQYLEKVELQEEBQVQREKRRERQRYLEK-----LQREBERL-----QEE 663
QY 197 AELEEHRSNQKMLQKKQSVQEKDHL-RGEHSKAVLARSKLESICRELQENRSL 255
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 664 EQLREERKRRERQRYLEKVELQEEBQLQREKREKQREKRRERYLEKBELOREERL 723
QY 256 KEEGVQBARBEERKREVTSHFQVTLNDIQLOMEQHN-ERNKSLRQENMELARLKKLIE 314
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 724 QREKQLQREDEKRRQVRERKYL---BEELQEEBRLQREKQLLREDEKQYLEKVEL 780
```

```
QY 315 QYELREHHIDKVEKHNDLQOLVDKLOQACQEMLKEAEERHOREKDFLLKEAVESORMCE 374
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
781 QRE--EEQLQREKRGRERQ-----YREBELREERLHRKEQQ-LORECECEKRROR 831
QY 375 LMKQOETHLKQOLALYTEKEFEF-----QNTLSKSSEVFTTPQEMERMTKIKKLEKE 428
Db | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
832 LERQLEEEELQRL----DRKRQFRDDDDQHQNEV-RNSRVVSKHRENKES----RQLD-D 881
QY 429 TTMYRSRWESSNKALLEMAFEKTVRDXELEGLQVK-----IQBLEKLCALQTERNDLNK 483
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
882 SWVRESQFQODLPLQDEQEBKEREHECEWMSRQKRDSPAPBQLLREEQOKETERDRKF 941
QY 484 RVQDLSAGGQ 493
Db | : :
942 REEEQLKGQ 951
```

Search completed: June 8, 2004, 16:35:10
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:27:01 ; Search time 17 Seconds
(without alignments)
1672.371 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 2785

Sequence: 1 MKNQDKKNGAKQSNPKSP.....AFSTASGQTGPQPTSPARA 546

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1315	47.2	259	YL14 HUMAN	P40222 homo sapien
2	267	9.6	1084	MYSS_RABIT	P02562 oryctolagus
3	263	9.4	2871	TRHY_HUMAN	P15924 homo sapien
4	261.5	9.4	1407	TRHY_RABIT	P37709 oryctolagus
5	260	9.3	886	RA50_ARCFU	O29230 archaeglob
6	259	9.3	1410	RRB1_HUMAN	Q9P269 homo sapien
7	257	9.2	1938	MYSS_CHICK	P13538 gallus gall
8	254.5	9.1	880	RA50_PVRAB	Q9U2C8 pyrococcus
9	253.5	9.1	1534	RRB1_CANFA	Q28298 canis faml
10	253.5	9.1	1549	TRHY_SHEEP	P22793 ovis aries
11	253.5	9.1	1690	C190_DROME	Q9V7E5 drosophila
12	249	8.9	1938	MYH4_RABIT	Q28541 oryctolagus
13	248	8.9	1605	RRB1_MOUSE	Q99P15 mus musculus
14	246.5	8.9	4473	PLEB1_CRIGR	Q9J155 cricetus
15	246	8.8	1427	REST_HUMAN	P30622 homo sapien
16	245	8.8	1927	MYH8_HUMAN	P13535 homo sapien
17	244.5	8.8	3259	GOB1_HUMAN	Q14789 homo sapien
18	243	8.7	1939	MYH1_HUMAN	P13533 homo sapien
19	243	8.7	1940	TRHY_CHICK	P02565 gallus gall
20	242.5	8.7	1898	TRHY_HUMAN	Q07283 homo sapien
21	242.5	8.7	4687	PLEB1_RAT	P30427 rattus norv
22	240.5	8.6	1934	MYH7_MSAU	P13540 mesocricetu
23	240	8.6	2663	CENE_HUMAN	Q02224 homo sapien
24	240	8.6	4684	PLEB1_HUMAN	Q15149 homo sapien
25	239.5	8.6	1939	MYH1_HUMAN	P12882 homo sapien
26	238.5	8.6	993	SCP1_MOUSE	Q62209 mus musculus
27	238	8.5	978	RA50_AQUAE	O67124 aquifex aeo
28	237	8.5	1411	EEAL_HUMAN	P15075 homo sapien
29	236	8.5	1935	MYH7_HUMAN	P12883 homo sapien
30	235	8.4	1938	MYH6_MOUSE	Q02566 mus musculus
31	235	8.4	1938	MYH6_RAT	P02563 rattus norv
32	235	8.4	1938	MYH6_HUMAN	Q9UKX3 homo sapien
33	234	8.4	1935	MYH7_PIG	P79293 sus scrofa

```

RESULT 1
YL14_HUMAN
ID YL14_HUMAN STANDARD; PRT; 259 AA.
AC P40222; OSN2Y3;
DT 01-FEB-1995 (Rel. 31, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein initially thought to be identical with
DE interleukin-14 (IL-14) (High molecular weight B-cell growth factor)
DE (HMW-BCGF).
GN IL14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93317675; PubMed=8327514;
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RT "Identification of a cDNA for a human high-molecular-weight B-cell
RT growth factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).
RN [2]
RP ERRATUM.
RX MEDLINE=96353961; PubMed=8755619;
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RL Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).
RN [3]
RP SEQUENCE OF 130-259 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton G., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shvchenko Y., Bouffard G.G.,
RA Blakesley A.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez R.C., Grooman J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002)
CC -!- CAUTION: The clone described in Ref.1 was initially thought to
CC code for a high molecular weight interleukin (IL-14). The
CC translation of this cDNA does not predict the open reading frame

```

ALIGNMENTS

CC for the 60 kDa protein described in Ref.1. A reading frame on the
 CC plus strand predicts a 7.7 kDa protein. The longest open reading
 CC frame (shown here) is on the opposite strand and predicts a 36.4
 CC kDa protein. The relationship of this sequence to IL-14, if any,
 CC is uncertain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; L15344; -; NOT ANNOTATED CDS.

DR EMBL; BC029686; AAH29686.1; -;

DR PIR; A48203; A48203.

DR Genew; HGNC:5976; IL14.

DR MIM; 147684; -;

DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0030372; F:high molecular weight B-cell growth factor . . .; NAS.

DR GO; GO:0008283; P:cell proliferation; NAS.

KW Hypothetical protein.

SQ SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;

Query Match 47.2%; Score 1315; DB 1; Length 259;

Best Local Similarity 100.0%; Pred. No. 2.7e-44;

Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 MEQHNERNKLRQENNELAKLKLIEQYELRESHIDKVPKHQIQQLVDAKLQQAQEM 347

Db 1 MEQHNERNKLRQENNELAKLKLIEQYELRESHIDKVPKHQIQQLVDAKLQQAQEM 60

QY 348 LKEAERHOREKDFLLKEAVESQRCMLKQOETHLKOALALYTKPEEFONTLSKSEV 407

Db 61 LKEAERHOREKDFLLKEAVESQRCMLKQOETHLKOALALYTKPEEFONTLSKSEV 120

QY 408 FTTFQENKEMTKKIKLEKETTWYRSWSSNKALLEWABEKTVRDKELGLQVKIQRL 467

Db 121 FTTFQENKEMTKKIKLEKETTWYRSWSSNKALLEWABEKTVRDKELGLQVKIQRL 180

QY 468 EKLCEALQTERNDLNKRVDLSAGQSLTDSGPRERPEGPGACQAPSSPRVTEAPCPYGA 527

Db 181 EKLCEALQTERNDLNKRVDLSAGQSLTDSGPRERPEGPGACQAPSSPRVTEAPCPYGA 240

QY 528 PSTEASGQTGPQEPSTASA 546

Db 241 PSTEASGQTGPQEPSTASA 259

RESULT 2

MYSS_RABIT

ID MYSS_RABIT STANDARD; PRT; 1084 AA.

AC P02562;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Myosin heavy chain, skeletal muscle (fragments).

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

FN (1)

RP SEQUENCE OF 1-258.

RA Capony J.-P., Elzinga M.;

RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of

RT myosin.";

RL Biophys. J. 33:148A-148A(1981).

SN (2)

RP SEQUENCE OF 259-428.

RX MEDLINE=85131142; PubMed=3972832;

RA Lu R.C.; Wong A.;

RT "The amino acid sequence and stability predictions of the hinge

RT

RL region in myosin subfragment 2.";

RN J. Biol. Chem. 260:3456-3461(1985).

RP (3)

RX SEQUENCE OF 409-1084 FROM N.A.

RA MEDLINE=87304245; PubMed=3305014;

RT Maeda K., Szakiel G., Wittinghofer A.;

RN "Characterization of cDNA coding for the complete light meromyosin

RL portion of a rabbit fast skeletal muscle myosin heavy chain.";

RX Eur. J. Biochem. 167:97-102(1987).

CC -!- FUNCTION: Muscle contraction.

CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2

CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)

CC and 2 regulatory light chain subunits (MLC-2).

CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing

CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,

CC characteristic for alpha-helical coiled coils.

CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light

CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be

CC split further into 2 globular subfragments (S1) and 1 rod-shaped

CC subfragment (S2). Contains 1 myosin-like globular head domain.

CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X05958; CAA23391.1; -;

DR PIR; A02985; A02985.

DR PIR; A05280; A05280.

DR PIR; S00084; S00084.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW ATP-binding; Multigene family.

FT NON TER 1

FT DOMAIN <1> 258 ALPHA-HELICAL TAIPICE (S2).

FT NON CONS 258 259

FT DOMAIN <259> 1084 RODLIKE TAIL (S2 AND LMM DOMAINS).

FT DOMAIN 455 1084 COILED COIL (POTENTIAL).

FT VARIANT 405 405 L -> V.

FT VARIANT 408 408 V -> L.

FT VARIANT 421 421 E -> D.

FT VARIANT 423 423 S -> G.

FT VARIANT 426 426 K -> R.

SQ SEQUENCE 1084 AA; 125488 MW; 229CFD69A6E1F7F0 CRC64;

Query Match 9.6%; Score 267; DB 1; Length 1084;

Best Local Similarity 22.9%; Pred. No. 0.00095;

Matches 137; Conservative 87; Mismatches 209; Indels 166; Gaps 21;

QY 4 QDKKNGAAKQSNPKSPGQPEAGPEGAQRPQAAPVAEAGPGSSQAPKPEGAQART 63

Db 172 EDKVNITLK---AKTKLEQVDDLESLEGEKKIRMDLE-----RAKELSG-DLKLA 220

QY 64 QSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQGEPAPEPAKSRVTVAR---- 119

Db 221 QETSM-DIENDKQQLDEKLKLEFMTNLQ-----SKIDEQALMTNLQRIEEL 267

QY 120 --NGEPEPTPVNGEKEPSKGDPTNTEIROS-DEVG-----DRHRRPQEKKAAGLG 169

Db 268 EEIEEAEASRAKAEKQSDLSRELEISERLEAGGASQAQIEMKKEAEFEKVERDL 327

QY 170 KEITLLMQTLNTLSTPEEKLAALCKKY----ABELLEHNSQOMKLLQKQSQLVQEKD 225

Db 328 EEATL-----QHEATAAALRKKHADSVBELGQIDNLQKVKOKLEKSEKSELAMEID 378

QY 226 HLRGEHSAKVLARSKLESLCKEL-----QSHNRSLSX----- 257

Db 379 DLAGNMETVSKAGNLEKMKRTLEDQLSEVTKTEESHQRLINLSAQAKRLHTESEFSR 438

QY	258	-----EGVQRAEEERKEKEVTSH-FQVTLNDIQLOMEQHNE	RA	Kowalczyk A.P., Bornslaeger E.A., Borgwardt J.E., Palka H.L.,
DB	439	QLEKDKAMVSQLSGGQAFQIQIEGLKRLQLEETKAKSALAHALQSSRDCLLRQYEE	RA	Dhalwal A.S., Corcoran C.M., Denning M.P., Green K.J.;
QY	294	R-----NSKLR-----QENMELAEELKLIQYELREZHIDKVFK	RT	"The amino-terminal domain of desmoplakin binds to plakoglobin and
DB	499	EQBAKAELOGRAMSKANSEVSWMTKCTDAIOETEELEAKKXLAORLODAEHV-----	RT	clusters desmosomal cadherin-plakoglobin complexes.";
QY	329	HKDLOQQLVDAKQQQAQEMLKAEERHQREKDFLLKEAVESQRCWELMFQOETHLQQLA	RL	J. Cell Biol. 139:773-784(1997).
DB	554	-----EAVNSKAS-----LEKTKQRLQNEAEDLMIDVERSNATCARMDKKQRPDKVLA	RN	[7]
QY	389	LYTEKEFEFQNTL-----SKSSEVFTTFK-----QEMVTKKIKLEKETNYS	RX	LIPIDATION.
DB	604	EWKHYTEQAELEASQKESRSLSSTEVFKVKNAYEESLDHLETKENKQLQOEISDLTE	RY	MEDLINE=98316349; PubMed=9651377;
QY	435	RWESSNKALLEMAEKTVRDKLEGLQVKIORLE-----KLCRALQTERNDLNKRV	RA	Marekov L.N., Steinert P.M.;
DB	564	QIAESAKHIELEKVKQIDQKSELSQALEEAGSLEHEEGKILR-IQLNLNQVSEI	RT	"Ceramide are bound to structural proteins of the human foreskin
			RT	epidermal cornified cell envelope.";
			RL	J. Biol. Chem. 273:17763-17770(1998).
			RN	[8]
			RX	VARIANT ARVD8 ARG-299.
			RY	MEDLINE=22285952; PubMed=12373648;
			RA	Rampazzo A., Nava A., Malacrida S., Beffagna G., Baucé B., Rossi V.,
			RA	Zimbello R., Simonati B., Basso C., Thiene G., Towbin J.A.,
			RA	Danieli G.A.;
			RT	"Mutation in human desmoplakin domain binding to plakoglobin causes a
			RT	dominant form of arrhythmogenic right ventricular cardiomyopathy.";
			RL	Am. J. Hum. Genet. 71:1200-1206(2002).
			RN	[9]
			RX	VARIANTS SPWHS LYS-287 AND CYS-2366.
			RY	MEDLINE=21830938; PubMed=11841538;
			RA	Whitlock M.V., Wan H., Morley S.M., Garzon M.C., Kristal L., Hyde P.,
			RA	McLean W.H.I., Pulkkinen L., Uitto J., Christiano A.M., Bady R.A.J.,
			RA	McGrath J.A.;
			RT	"Compound heterozygosity for non-sense and mis-sense mutations in
			RT	desmoplakin underlies skin fragility/woolly hair syndrome.";
			RL	J. Invest. Dermatol. 118:232-238(2002).
			CC	-I- FUNCTION: Major high molecular weight protein of desmosomes.
			CC	Involved in the organization of the desmosomal cadherin-
			CC	plakoglobin complexes into discrete plasma membrane domains and in
			CC	the anchoring of intermediate filaments to the desmosomes.
			CC	-I- SUBUNIT: Homodimer.
			CC	-I- SUBCELLULAR LOCATION: Innermost portion of the desmosomal plaque.
			CC	-I- ALTERNATIVE PRODUCTS:
			CC	Event-Alternative splicing; Named isoforms=2;
			CC	Name=DPI; Synonyms=DPI1;
			CC	IsoId=P15924-1; Sequence=Displayed;
			CC	Name=DPI1; Synonyms=DP2;
			CC	IsoId=P15924-2; Sequence=VSP_005070;
			CC	-I- TISSUE SPECIFICITY: Isoform DPI is apparently an obligate
			CC	constituent of all desmosomes; Isoform DPI1 resides predominantly
			CC	in tissues and cells of stratified origin.
			CC	-I- DOMAIN: The N-terminal region is required for localization to the
			CC	desmosomal plaque and interacts with the N-terminal region of
			CC	plakophilin 1. The C-terminal region interacts with intermediate
			CC	filaments.
			CC	-I- PTM: Substrate of transglutaminase. Some glutamines and lysines
			CC	are cross-linked to other desmoplakin molecules, to other proteins
			CC	such as keratin, envoplakin, periplakin and involucrin, and to
			CC	lipids like omega-hydroxyceramide.
			CC	-I- DISEASE: Defects in DSP are the cause of familial arrhythmogenic
			CC	right ventricular dysplasia-8 (ARVD8) [MTM:607450], an autosomal
			CC	dominant disorder.
			CC	-I- DISEASE: Defects in DSP are a cause of striate palmoplantar
			CC	keratoderma II (PPKS2, KPPS2 or SPPK2), characterized by skin
			CC	thickening in the palms (linear pattern) and the soles (island-
			CC	like pattern) and flexor aspect of the fingers; and rarely by
			CC	abnormalities of the nails, the teeth and the hair.
			CC	-I- DISEASE: Defects in DSP are the cause of skin fragility-woolly
			CC	hair syndrome (SPWHS) [MIM:607655], an autosomal recessive
			CC	genodermatosis characterized by focal and diffuse palmoplantar
			CC	keratoderma, hyperkeratotic plaques on the trunk and limbs, and
			CC	woolly hair with varying degrees of alopecia.
			CC	-I- SIMILARITY: Contains 17 plectrin repeats.
			CC	-I- SIMILARITY: Contains 2 spectrin repeats.
			CC	-I- SIMILARITY: Belongs to the plakin or cytolinker family.
			CC	-----
			CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
			CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
			CC	the European Bioinformatics Institute. There are no restrictions on its

CCC use by non-profit institutions as long as its content is in no way
CCC modified and this statement is not removed. Usage by and for commercial
CCC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CCC or send an email to license@isb-sib.ch).

DR	EMBL; M77830; AAB85135.1; --	
DR	EMBL; AL031058; --; NOT_ANNOTATED_CDS.	
DR	EMBL; J05211; AAB35786.1; --	
DR	EMBL; AF139065; AAF19785.1; --	
DR	PIR; A38194; A38194.	
DR	PDB; 1LM5; 31-JUL-02.	
DR	PDB; 1LM7; 31-JUL-02.	
DR	Genew; HGNC:3052; DSP.	
DR	MIM; 125647; --	
DR	MIM; 607450; --	
DR	MIM; 607655; --	
DR	GO; GO:0005200; F:Structural constituent of cytoskeleton; TAS.	
DR	GO; GO:0008544; P:Epidermal differentiation; TAS.	
DR	InterPro; IPR001101; Plectin repeat.	
DR	InterPro; IPR002017; Spectrin.	
DR	Refam; PF00681; Plectin; 11.	
DR	SMART; SM00250; PLEC; 17.	
DR	Repeat; Coiled coil; Phosphorylation; Cytoskeleton;	
KW	Structural protein; Alternative splicing; Disease mutation;	
KW	3D-structure; Lipoprotein.	
KW	DOMAIN 1 1056	GLOBULAR 1.
FT	DOMAIN 1057 1945	CENTRAL FIBROUS ROD DOMAIN.
FT	DOMAIN 1946 2871	GLOBULAR 2.
FT	REPEAT 347 447	SPECTRIN 1.
FT	REPEAT 858 952	SPECTRIN 2.
FT	DOMAIN 1018 1945	COILED COIL (POTENTIAL).
FT	REPEAT 2009 2045	PLECTIN 1.
FT	REPEAT 2046 2083	PLECTIN 2.
FT	REPEAT 2084 2121	PLECTIN 3.
FT	REPEAT 2122 2159	PLECTIN 4.
FT	REPEAT 2163 2197	PLECTIN 5.
FT	REPEAT 2198 2233	PLECTIN 6.
FT	REPEAT 2251 2288	PLECTIN 7.
FT	REPEAT 2289 2326	PLECTIN 8.
FT	REPEAT 2327 2364	PLECTIN 9.
FT	REPEAT 2365 2402	PLECTIN 10.
FT	REPEAT 2406 2440	PLECTIN 11.
FT	REPEAT 2456 2493	PLECTIN 12.
FT	REPEAT 2507 2544	PLECTIN 13.
FT	REPEAT 2610 2647	PLECTIN 14.
FT	REPEAT 2648 2685	PLECTIN 15.
FT	REPEAT 2724 2761	PLECTIN 16.
FT	REPEAT 2762 2799	PLECTIN 17.
FT	DOMAIN 2824 2847	6 x 4 AA TANDEM REPEATS OF G-S-R-[SR].
FT	DOMAIN 1 584	INTERACTS WITH PLAKOPHILIN 1 AND JUNCTION PLAKOGLOBIN.
FT	LIPID 2480 2480	Omega-hydroxyceramide glutamate ester (Potential).
FT	MOD_RES 2849 2849	PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT	VAR_SPLIC 1195 1794	Missing (in isoform DFI).
FT		/FTid=VSP_005070.
FT	VARIANT 287 287	N -> K (in SFWHS).
FT		/FTid=VAR_015569.
FT	VARIANT 299 299	S -> R (in ARVD8).
FT		/FTid=VAR_015402.
FT	VARIANT 2366 2366	R -> C (in SFWHS).
FT		/FTid=VAR_015570.
FT		R -> A (in REF. 2).
FT	CONFLICT 905 905	D -> R (in REF. 3).
FT	CONFLICT 1120 1120	SV -> RL (in REF. 2).
FT	CONFLICT 2687 2688	
FT	SEQUENCE 2871 AA; 331771 MW; 5639B5B7CD4690B7 CRC64;	

```
Query Match      9.4%; Score 263; DB 1; Length 2873;
Best Local Similarity 23.4%; Pred. No. 0.0034;
Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 25;

Qy      42 EAEKPGSSQAPRKP-----EGAAQAATAG-----GAERDVSELSERQLEDIL 83
```

Qy 42 EAEGPGSSQAPRKP-----EQAQAPTAQS-----GALRDVSELSRQLEDIL 83
 . . | | | | : | | | : : : : : : : :
 : : : : : : : : : : : : : : : : : :

[illegible]

```

RESULT 4
TRHY_RABIT
ID TRHY_RABIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trichohyalin.
TH TH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Stheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_Taxid=9986;
[1]
SEQUENCE FROM N.A.
RA Pietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular
CC layer of the epidermis. It later becomes cross-linked to KIF by
CC isodipeptide bonds. It may serve as scaffold protein, together
CC with involucrin, in the organization of the cell envelope or even
CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent post-synthetic processing during terminal
CC differentiation.
CC -1- SUBUNIT: Homodimer (Probable).
CC -1- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -1- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely

```

CC alpha-helical, configured as a series of peptide repeats of
CC varying regularity, and are thought to form a single-stranded
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC the most regular and may bind KIP directly by ionic interactions.
CC Domains 5 and 7 are less well organized and may induce folds in
CC the molecule. Domain 9 contains the C-terminus, conserved among
CC different species.
CC
CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are
CC probably converted to citrullines by peptidylarginine deiminase.
CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100
CC family.
CC
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z19092; CAA79519.1; -.
CC PIR; S28589; S28589.
CC HSP; P02633; 41CB.
CC InterPro; IPR001751; GABP_S100.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; EF-hand_1.
CC Pfam; PF01023; S_100_1.
CC ProDom; PD003407; GABP_S100; 1.
CC PROSITE; PS00018; EF_HAND; 1.
CC PROSITE; PS00303; S100_CABP; 1.
CC Keratinization; Calcium-binding; Repeat; Citrullination.
CC DOMAIN 1 91 S-100 LIKE.
CC CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
CC FT CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
CC SEQUENCE 1407 AA; 103781 MW; A617D2A159F12B7P CRC64;
Query Match 9.4%; Score 261.5; DB 1; Length 1407;
Best Local Similarity 23.5%; Pred. No. 0.002;
Matches 117; Conservative 110; Mismatches 168; Indels 103; Gaps 22;
QY 42 BAEGPGSS-----QAPRPEGAQARTAGSALRDVSEILSLQLELDILSTCYVDNQQG 94
DB 514 QEQPGQWNLQEEAQRHHTUYAPGQOEURE--EEELQR-----EKQRQE 561
QY 95 PGEDAGCEPAEPEDABEDASRTYVARGPEPTPVNGKEPKSGDNPTEBIROSDVGR 154
DB 562 REREVEEEKLQREDEKER-----RQERQYRELESLRQEEQLDR 604
QY 155 DHRPQKKKANGKGGKIKITLMQTLNTLSTPEKLAALCKKYAELLE--HNKQKQMKLL 213
DB 605 KLREEEQLQER---EERLRQRERKLRREEQL--LRQEEQLRQERERKLRREEQL 659
QY 214 QKKSQVLQVQKH--LRGHSKAVLRKSLSLRELQHNRSLEKGVQVQARSEE--EK 269
DB 560 RREERLQERERKLRREEQ---LLQREERLRRQERARKLRREEQLLRQEEQLRQER 716
QY 270 RKEVTSFQVTLNDIQOMQHNRNSKLQENNEL-----AERLKKLIEQYELREHIDK 325
DB 717 ERKLREEEQLLRQEEQLRQ---ERDRKLREEEQLQEESEERLRQERQQLRRER--DR 772
QY 326 VFHKD-----LQQQLVDKLAQAQEMLKEE-----ERHOREKDFLLKEA 366
DB 773 KPREEQQLQREERLRQERERKLRREEQLLRQEEQLRQERERKLRREEQLRQER 832
QY 367 VESQMC-----ELMKQOETHLQKQALYTKFEFEFQNTLSKSEVFTTFKQE 414
DB 833 -EERLRQERERKLRREEQLLRQEEQLRQERA---RKLREEEQLRQEEQ---HLRQE 885
QY 415 MEKMTKKIKKL-----EKETTMYRSEWSSKALLEMAEKTVDKLEGLQVKIQRLK 470
DB 886 RDRKLREEEQLRQEEQLRQERDRKLRREEQLLRQEESEERLRQERER---KLREEEQL 942

QY 471 CRALQTE-RNDLNKRVQD 487
DB 943 LRREEQELRRERARKLRE 960
RESULT 5
ID RA50_ARCFU STANDARD; PRT; 886 AA.
AC Q29230;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR AF1032
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-15 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.W.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A3001032; AAB90211.1; -.
DR PIR; H69378; H69378.
DR TIGR; AF1032;
DR HAMAP; MF_0049; 1.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR007517; Rad50_zn_hook.
DR InterPro; IPR003395; SMC N.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR Pfam; PF02463; SMC_N; 1.
DR SMART; SM00382; AAA; 1.
DR DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 31 38 ATP (BY SIMILARITY).
FT DOMAIN 148 728 COILED COIL (POTENTIAL).
SQ SEQUENCE 886 AA; 103633 MW; D35641D499AAB58 CRC64;
Query Match 9.3%; Score 260; DB 1; Length 886;
Best Local Similarity 24.6%; Pred. No. 0.0015;
Matches 118; Conservative 90; Mismatches 118; Indels 154; Gaps 23;

QY 108 EDAKSRVTVARNGEPPTPVNGEKEPSKGDPTNTEERQSDVEVDHRRPQPKKAG 167
 DB 155 ENAMKNGAVIR-----MLREKERLK-----EFLSQBEQI-----KKQKBEKA-- 194
 QY 168 LGKEITLMTQNTLSTPEKLAALCKKYAEELLEHRNSQKMLLKQKQKQKQKQK 227
 DB 195 ---ETIERSEBIKXIESLREKLS-----BEVRNLESLRKLKLEHKSRL-----ESL 237
 QY 228 RGEHSAVLAKSLKSLRELOHNRSLKEBGVQBARBEKKEV----- 273
 DB 238 RKQESSVLCEVRGLEKLELE---KQKEV-VERIEDLEKAKVEKELKPAERYSILE 293
 QY 274 ---TSHFQVTLND-----IQLOEQHNRNSKLQENMELAEKLLKLEOYE 317
 DB 294 KLSLEINQALRDVEKRGDLTEAAGIOALKKAEEDNSKL-ETIKTKIBELERELRFE 352
 QY 318 -----LREHI---DKVFKHLOLOOLVDKALQOQAEMLKAAE 353
 DB 353 KSHRLLETLPKMDRMQGIKALBEKNLTPDKVKMYDL-----LSKAKBEKEITE 404
 QY 354 RHQR-----LKEAVE-----SORMCELMKQO---ETHLKOOLALYTEKFBFQ 398
 DB 405 KKLKLIKSSIKTGAQKAKVAELKSAERTCPVCGRELDEKRNIMAYITREMRIA 464
 QY 399 NTLKSSSVFTTFOEMERMTKIKLKEKTTMVRSW-----ESSNKALL 444
 DB 465 EELAKADIEIKLERLEKVE---KALEKQETVLKRYQWDELKALENELSHDAEKL 521
 QY 445 EMAREKTVRDKELEGL---OVKI-----ORLEKLCALQOTRENDLANKVQD 487
 DB 522 ESEYRKVKER-LOGLRQOQKILLSSASRIKELKSLREIBALKNVESEGEHLRKIRE 580

RESULT 6
 RRB1_HUMAN
 ID RRB1_HUMAN STANDARD; PRT: 1410 AA.
 AC Q9P2E9; Q75300; Q75301; Q96SB2; Q9BWP1; Q9HA76;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DE Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa
 ribosome receptor homolog) (ES/130 related protein).
 GN RRB1 OR KIA1398.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=98290552; PubMed=9628588;
 RA Langley R., Leung E., Morris C., Berg R., McDonald M., Weaver A.,
 RA Parry D., Ni J., Su J., Gentz R., Spurr N., Krissansen G.W.;
 RT "Identification of multiple forms of 180-kDa ribosome receptor in
 human cells.";
 RL DNA Cell Biol. 17:449-460(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leheslahti M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLeay K., McMurray A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [4]
 RP SEQUENCE OF 413-1208 FROM N.A. (ISOFORM 3).
 RC TISSUE=Uterus;
 RA Wabbutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 756-1410 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
 RA Bosak S.A., McEwan P.J., Kerneran K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallal D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Acts as a ribosome receptor and mediates interaction
 between the ribosome and the endoplasmic reticulum membrane (By
 similarity).
 CC -!- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
 reticulum (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=3;
 Comment=Additional isoforms seem to exist. Additional isoforms
 are probable deriving from alternative splicing in the repeat
 region;
 CC Name=3;
 CC IsoId=Q9P2E9-1; Sequence=Displayed;
 CC Name=1; Synonyms=pi180;
 CC IsoId=Q9P2E9-2; Sequence=VSP_003950;
 CC Name=2; Synonyms=ES130;
 CC IsoId=Q9P2E9-3; Sequence=VSP_003949, VSP_003950;
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

SEQUENCE OF 838-1938.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041770; PubMed=1939030;
RA Maita T., Tajima E., Nagata S., Miyaniishi T., Nakayama S., Matsuda G.,
RT "The primary structure of skeletal muscle myosin heavy chain: IV.
RT Sequence of the rod, and the complete 1,938-residue sequence of the
RT heavy chain.";
RL J. Biochem. 110:75-87(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 1-808.
RX MEDLINE=87092420; PubMed=3467365;
RA Maita T., Hayashida M., Tanioka Y., Komine Y., Matsuda G.,
RT "The primary structure of the myosin head.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:416-420(1987).
RN [7]
RP SEQUENCE OF 842-1270.
RX MEDLINE=90121764; PubMed=2610940;
RA Watanabe B.,
RT "Complete amino-acid sequence of subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:1027-1034(1989).
RN [8]
RP SEQUENCE OF 852-1108.
RX MEDLINE=89374803; PubMed=2775482;
RA Watanabe B.,
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:549-558(1989).
RN [9]
RP SEQUENCE OF 1145-1270.
RX MEDLINE=89228549; PubMed=2713098;
RA Watanabe B.,
RT "Amino-acid sequence of the hinge region in chicken myosin
RT subfragment-2.";
RL Biol. Chem. Hoppe-Seyler 370:55-61(1989).
RN [10]
RP SEQUENCE OF 1857-1938 FROM N.A.
RX MEDLINE=87217964; PubMed=3034534;
RA Moriarty D.M., Barringer K.J., Dodgson J.B., Richter H.B.,
RT Young R.B.,
RT "Genomic clones encoding chicken myosin heavy-chain genes.";
RL DNA 6:91-99(1987).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-843.
RX MEDLINE=93303624; PubMed=8316857;
RA Rayment I., Rypniewski W.R., Schmidt-Base K., Smith R.,
RA Tomchick D.R., Benning M.M., Winkelmann D.A., Wesenberg G.,
RT "Three-dimensional structure of myosin subfragment-1: a molecular
RT motor.";
RL Science 261:50-58(1993).
CC -!- FUNCTION: Muscle contraction. Myosin is a protein that binds to
CC F-actin and has ATPase activity that is activated by F-actin.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; U87231; AAB47555.1; -;
CC EMBL; M16557; AAA48970.1; -;
DR PDB; 2MYS; 11-JAN-97.
DR PDB; 1ALM; 17-DEC-97.
DR PDB; 1M8Q; 10-SEP-02.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001603; Myosin head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00736; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
DR MYOSIN; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Phosphorylation; Acetylation;
KW Calmodulin-binding; Multigene family; 3D-structure.
FT INIT MET 0 0 MYOSIN HEAD-LIKE.
FT DOMAIN 1 782 IQ.
FT DOMAIN 783 812 HINGE.
FT DOMAIN 838 840 COILED COIL (POTENTIAL).
FT DOMAIN 841 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 35 35 METHYLATION (MONO-).
FT MOD_RES 130 130 METHYLATION (TRI-).
FT MOD_RES 551 551 METHYLATION (TRI-).
FT MOD_RES 755 755 METHYLATION (MONO-).
FT CONFLICT 907 907 C -> Q (IN REF. 7 AND 8).
FT CONFLICT 980 980 L -> F (IN REF. 1).
FT CONFLICT 1343 1343 E -> D (IN REF. 5).
FT CONFLICT 1545 1545 S -> A (IN REF. 5).
FT CONFLICT 1796 1797 HV -> QL (IN REF. 5).
FT CONFLICT 1830 1830 S -> A (IN REF. 5).
FT CONFLICT 1863 1863 I -> V (IN REF. 10).
FT CONFLICT 1929 1931 IHG -> FH (IN REF. 10).
FT TURN 7 8
FT TURN 11 12
FT TURN 14 17
FT TURN 21 24
FT TURN 25 28
FT TURN 34 36
FT TURN 37 41
FT STRAND 47 55
FT STRAND 59 62
FT STRAND 70 73
FT STRAND 74 76
FT STRAND 77 78
FT TURN 83 84
FT STRAND 90 90
FT STRAND 91 93
FT HELIX 99 108
FT TURN 109 113
FT STRAND 116 118
FT STRAND 123 126
FT STRAND 133 135
FT TURN 137 138
FT TURN 139 142
FT TURN 143 145
FT TURN 148 149
FT TURN 155 169
FT TURN 170 170
FT TURN 173 179
FT TURN 181 182
FT TURN 185 185
FT TURN 185 185


```
QY 401 LSKSSEVPTTF-KQEMKMTKIKKLEKTTMYSSRMESSNKA-----LLMAREKTV 452
Db 647 EAKLUDSESEFESEKREKLVLEKREVSLLTLEBKKSVEQIKATLRKLEKEE 706
QY 453 RDEKEGLQVKIQRLEKLCRALQOTERNDLNKRVOD 487
Db 707 REK-----AKLEIKKLEKALSKEVE-----DLRKKIND 733

RESULT 9
RRB1 CANFA
ID -RRB1 CANFA STANDARD; PRT; 1534 AA.
AC Q28298;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome-binding protein 1 (180 kDa ribosome receptor) (RRP).
GN RRBP1 OR P180.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Kidney;
RX MEDLINE=95310363; PubMed=7790375;
RA Wanker E.E., Sun Y., Savitz A.J., Meyer D.I.;
RT "Functional characterization of the 180 kDa ribosome receptor in
RT vivo."
RL J. Cell Biol. 130:29-39(1995).
CC -!- FUNCTION: Acts as a ribosome receptor and mediates interaction
CC between the ribosome and the endoplasmic reticulum membrane.
CC -!- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
CC reticulum.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk.)
CC -----
CC EMBL; X87224; CAA60676.1; -.
DR FIR; A56734; A56734.
DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IDA.
DR GO; GO:0030376; F:ribosome receptor activity; IDA.
DR GO; GO:0007165; P:signal transduction; IDA.
DR InterPro; IPR007794; Rib recp KP reg.
DR Pfam; PF05104; Rib recp KP reg. I.
KW Transport; Protein transport; Translocation; Endoplasmic reticulum;
KW Transmembrane; Repeat; Alternative splicing.
FT DOMAIN 1 LUMENAL (POTENTIAL).
FT TRANSMEM 8 28 POTENTIAL.
FT DOMAIN 29 1534 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 197 736 54 X 10 AA APPROXIMATE TANDEM REPEATS OF
FT N-Q-G-K-K-A-E-G-A-P.
FT SEQUENCE 1534 AA; 164586 MW; B343BCF12656F3C5 CRC64;

Query Match 9.1%; Score 253.5; DB 1; Length 1534;
Best Local Similarity 20.7%; Pred. No. 0.0043;
Matches 158; Conservative 115; Mismatches 229; Indels 261; Gaps 31;

QY 3 NODKNGGAQKQSNPKSSPGQEPAGPE-----GAQERPSQAA 38
Db 697 NQKK-GEFTNQKSGSESPNQKGVASANQSKRASAPIQGNADMVQSQEAPKQEA 755
QY 39 PAVEAEGPGSGQAPKRP-----EQAQAR-----TAQSGALRD 70
Db 756 PAKKISGKKGKGPFGPDSPLYPKTLTVSGVMVFNQGEAQRLEILSEKAGVTQD 815
QY 71 VSEB-----LSRQLED-----ILSTVCVDNNQGGPGEDGAQGEPAERPDAAKSR-- 115
```

```
816 TWHKATQKGDPAVAILKRLQLEKEKLLAT-----EQEADAQVAKSKL 855
116 -VYARNGSEPTFPVNGSEKPSK-----GDPNTEETIQSDVGDHRRRPOEKKKAK--- 166
856 REVNKELAAKAKAAGEAKVKKQVLVAREQETAVQARIEASYREHVKEVQQLQGIKRTL 915
167 -----GLGKEITLLMOTLN-----TLSTPEKLAALCKKYAELE 201
916 QEQLNGPNTQLARLQOENSILRDALNQATSVQESQNTLAKLROELSKVSKELVSE 975
202 EHRNSQKMLKQKQKQSVQKDHGRGHSKAVLA-RSKLSLQREL-----QRHNSLKE 257
976 AARQEQQRKALETKTAALKQVQLQASHKSEELQRLDEVSELRCSTSHASLRA 1035
258 EGVQARAREEKKRKEVTSHFQVTLNDIQLQMEQHNRNSKL-----ROENMELAEKLC--- 311
1036 D-AEKAQEQQQQVQELHSLKQSSAEVSKSELSGLHGQLKEARASQLMERIRSIEA 1094
312 LIOYELRE-----EHIDKVFHKDQLQQLVDKLAQQAQEMLKKEERHQ----- 356
1095 LLEAGQARDTQDAQASRAEHQARL-----KELESQV--WCLEKEATELKEAVEQKRVKNDL 1149
357 REKDFLLKEAVES-QRMCELMKQOETHLQKQ-----LAL----- 389
1150 REKNWAMEALASAEACBESKLSLQAKESEKQSLTEAQTKKALLALLPALSSAPQ 1209
390 -YTE-----KFEFQNTLSKSSSEVFTFKQEMEK 417
1210 SYTENLQELREKGPPELLKQRPADTPDPSDLASKLREABETQNNLQACQYRTILATEG 1269
418 MTKKIKK-LEKETTMYRSRWSNKKALLEMAEKTVRDKLEGLQVKIQ----- 465
1270 MLKDLQKSEVEBQV-----WKAKVSATEBELQKSRVTVKHELDI VEKLGELSSQVRE 1325
466 -----RIEKCGRALQTRNDLNKRVQDLSAGQGSGLTDSGPERPEGCAQAPS--- 515
1326 HTSHLEAELEKHNMAASAECCQSYAKEV-----AGLRQLLESQSLDAAKSEAQKSNELA 1381
516 ----PRYTEAPCY-----PGAPSTEASGQTGPQBPSTARA 546
1382 LVYQQLSEKMSKSHVEDGDVAGSPAAPAE-----QDPVELKA 1417

RESULT 10
TRHY SHEEP
ID -TRHY SHEEP STANDARD; PRT; 1549 AA.
AC P22793;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trichohyalin.
GN THH.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93260018; PubMed=7684041;
RA Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;
RT "Analysis of the sheep trichohyalin gene: potential structural and
RT calcium-binding roles of trichohyalin in the hair follicle."
RL J. Cell Biol. 121:855-865(1993).
RN [2]
RP SEQUENCE OF 1016-1549 FROM N.A.
RC STEAIN-Merino-Dorset horn X Border Leicester; TISSUE=Wool follicles;
RX MEDLINE=90130632; PubMed=2298812;
RA Fietz M.J., Presland R.B., Rogers G.E.;
RT "The cDNA-decoded amino acid sequence for trichohyalin, a
RT differentiation marker in the hair follicle, contains a 23 amino acid
RT repeat.";
```


RESULT 11
 C190_DROME STANDARD; Q8INY8; Q8MSD0; PRT; 1690 AA.
 AC Q9VUE5; Q44929; Q8INY8; Q8MSD0;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Restin homolog (Cytoplasmic linker protein 190) (Microtubule binding
 protein 190) (d-CLIP-190).
 GN CLIP-190 OR Q5020.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A), FUNCTION, SUBCELLULAR LOCATION, AND
 RP TISSUE SPECIFICITY.
 RC STRAIN=Oregon-R; TISSUE=Embryo, and Ovary;
 RX MEDLINE=98139549; PubMed=9472041;
 RA Lantz V.A., Miller K.G.;
 RT "A class VI unconventional myosin is associated with a homologue of a
 RT microtubule-binding protein, cytoplasmic linker protein-170, in
 RT neurons and at the posterior pole of Drosophila embryos.";
 RL J. Cell Biol. 140:897-910(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
 RA Wan K.H., Doyle C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hooten D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Martell B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Murgall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC STRAIN=Berkely; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarini H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 RN [5]
 RP FUNCTION, SUBUNIT, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RC MEDLINE=20530668; PubMed=11076973;
 RA Sisson J.C., Field C., Ventura R., Royou A., Sullivan W.;
 RT "Java lamp, a novel peripheral Golgi protein, is required for
 RT Drosophila melanogaster cellularization.";
 RL J. Cell Biol. 151:905-918(2000).
 CC -1- FUNCTION: Together CLIP-190 and jar may coordinate the interaction
 CC between the actin and microtubule cytoskeleton. May link endocytic
 CC vesicles to microtubules. May play a role in formation of furrows
 CC during cellularization.
 CC -1- SUBUNIT: Interacts with Lva.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; microtubule-associated. Lva-
 CC CLIP-190 complexes are found at the Golgi.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Name=A;
 CC Name=B;
 CC Name=C;
 CC Note=No experimental confirmation available;
 CC IsoId=Q9VUE5-2; Sequence=VSP_050479;
 CC Name=C;
 CC Note=No experimental confirmation available;
 CC IsoId=Q9VUE5-3; Sequence=VSP_050480;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Specifically expressed at the tip of the
 CC furrow in cellularizing blastoderm. CLIP-190 and jar are
 CC coexpressed at several times in development and in a number of
 CC tissues, including embryonic axonal neuron processes and posterior
 CC pole.
 CC -1- SIMILARITY: Contains 2 CAP-Gly domains.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EXBL; AF041382; AAB96783.1; -
 CC EXBL; AB003655; AAF53604.1; -
 CC EXBL; AB003655; AAF53605.2; -
 CC EXBL; AB003655; AAN10987.1; -
 CC EXBL; AY118896; AAM50756.1; -
 CC Flybase; Fgmd020503; CLIP-190.
 CC GO; GO:0005938; C:cell cortex; IDA.
 CC GO; GO:0005794; C:Golgi apparatus; IDA.
 CC GO; GO:0005794; C:Golgi apparatus associated complex; IDA.
 CC GO; GO:0003779; F:actin binding; IDA.
 CC GO; GO:0008017; F:microtubule binding; IDA.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC GO; GO:0007349; P:cellularization; IMP.
 CC InterPro; IPR000936; CAP-Gly.
 CC Pfam; PF01302; CAP GLY; 2.
 CC PROSITE; PS00845; CAP GLY 1; FALSE_NEG.
 CC PROSITE; PS02445; CAP_GLY_2; 2.


```

FT DOMAIN 784 813 IQ.
FT DOMAIN 842 1938
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A9A2EC5B182626 CRC64;

Query Match
Best Local Similarity 8.9%; Score 249; DB 1; Length 1938;
Matches 113; Conservative 80; Mismatches 179; Indels 160; Gaps 16;

QY 49 SQAKPKPGQAQTAQSGALDVSEELSRQLEDILSTVCVNNQGGPGDGAQGEPAPE 108
DB 1091 SNLSKIDQALAMQ--LQKKIQLQARTEEL-----EEIEIERASRA 1133
QY 109 DAEKSTVAVANGBPEPTFVNVGKEPKSGDPNTEIRQS--DEVDGDRHRRPQEKKAKG 167
DB 1134 KAEKQSDLSR-----ELIEISERLEAGGATSAGIEMNKKREA 1172
QY 168 LGKETTLMOQLNTLSTPEKLAALCKK-----YAELEBHRNSQKMLQKXOSOLVQ 222
DB 1173 ---EFQKMRDLERATLQHEATAATLRKKHADSVAEELGEGQIDNLQRVKQKLEKEKSELKM 1229
QY 223 EKHLGRGSKAVLARSLKSLCREL-----QRNRSLEKGVQEAR----- 264
DB 1230 EIDDLASNNETVSKAGNLEKWCRTLEDQVSELTKKEEHRLLINDLSAQARLQTSGE 1289
QY 265 -----BEEKREKVTSH-FQVTLNDIQLQMEQ 290
DB 1290 FSRQLDEKSLVSLSRGKAFTQIEELKQLESEETKAKSALAHQSAHRDCDLLREQ 1349
QY 291 HNER-----NSKL-----QENNELAERLKKLEQVELEEEHDX 325
DB 1350 YEEQRAKAEIQRAMSKANSEVAQMTKYETDAIQRTEELSEAKKLAQLRQDAEEHV-- 1407
QY 326 VFKHDLQQLVDKLAQAQAEERHOREKDFLLKEAVESQRCMLAKQOETHLKQ 385
DB 1408 -----EAVNAKAS-----LEKTKQRLQNEVEDLMIDVETNAACAALDKQRPDK 1454
QY 386 QIALYTEKEFPQNTLSKSSVFYTFKQEMEMTKKIKLEKETVYRSRWESNKALL 445
DB 1455 ILAEKWKHYETH-----AELEASQKESGSLSTEVFKVKNAYEESLDQLET 1500
QY 446 MAEKTVRDKLEGL--QV-----KIQRLKLCALQTERNDLNKRVQDLQA 490
DB 1501 LKREKNEQQHISDLTEQIAEGGKRIHELEKVKQVEQKSELQAALAEARA 1552

RESULT 13
RRB1 MOUSE
ID_RRB1 MOUSE STANDARD; PRT; 1605 AA.
AC Q99PL5; Q99PK5; Q99PK6; Q99PK7; Q99PK8; Q99PK9; Q99PL0; Q99PL1;
AC Q99PL2; Q99PL3; Q99PL4; Q9CS20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome-binding protein 1 (Ribosome receptor protein) (mRrp).
GN RRPp1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP RRP15A; RRP15B; RRP10; RRP5.4; RRP2; RRP1.8 AND RRP0; AND TISSUE
RP SPECIFICITY.
RC STRAIN=ICR; TISSUE=Embryo;
RX MEDLINE=21104429; PubMed=11167022;
RA Kim Y.-J., Lee M.-C., Kim S.-U., Chun J.-Y.;

```

```

*Identification and characterization of multiple isoforms of a mouse
ribosome receptor.*;
Gene 261:337-344 (2000).
[2]
SEQUENCE FROM N.A. (ISOPORM 2).
TISSUE=Kidney;
MEDLINE=22388257; PubMed=12477932;
Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
Klauser R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Browstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
Rosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Vallalath D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Whiting J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
*Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.*;
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[3]
SEQUENCE OF 736-1605 FROM N.A. (ISOPORM 3).
STRAIN=C57BL/6J; TISSUE=Embryo;
MEDLINE=21085660; PubMed=11217851;
Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
Kadohira K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Browstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
Havashizaki Y.;
*Functional annotation of a full-length mouse cDNA collection.*;
Nature 409:885-890 (2001).
-i- FUNCTION: Acts as a ribosome receptor and mediates interaction
between the ribosome and the endoplasmic reticulum membrane (by
similarity).
-i- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
reticulum (by similarity).
-i- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=13;
Comment=Additional isoforms seem to exist. Experimental
confirmation may be lacking for some isoforms;
Name=3;
IsoId=Q99PL5-1; Sequence=Displayed;
Name=RRP61; Synonym=psp180;
IsoId=Q99PL5-2; Sequence=VSP_003963, VSP_003964;
Name=RRP47; Sequence=VSP_003959, VSP_003963, VSP_003964;
IsoId=Q99PL5-3; Sequence=VSP_003961, VSP_003963, VSP_003964;
Name=RRP41;
IsoId=Q99PL5-4; Sequence=VSP_003954, VSP_003962, VSP_003963;
Name=RRP16.8;
IsoId=Q99PL5-5; Sequence=VSP_003954, VSP_003964, VSP_003963;
Name=RRP15a;
IsoId=Q99PL5-6; Sequence=VSP_003958, VSP_003963, VSP_003964;

```



```
QY 422 IKKLEKETTMYRSESSNKALLEMEEKTVRDKLEGLQVKIQRLEKLCRALQTERNDL 481
DB 2464 QRMQEQKELVASMEEARRRQCEAEAVRQKEELQHLELQFQQCEKL--LAEEVQRL 2520
QY 482 NKEVDLSAGQCSLYSDGPERPEPGQAQ-PSSPRVTEAPCYCAPSTEASG 534
DB 2521 RERLQREBEHRAALAHSEBIAATAAAKALPGRDALDGPMSVEVEPEHAFEG 2574

RESULT 15
REST_HUMAN STANDARD; PRT; 1427 AA.
AC F30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
  Sternberg intermediate filament associated protein).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood monocytes;
RX MEDLINE=92289675; PubMed=1600942;
RA Bilbe G., Delaite J., Bruegggen J., Richener H., Asselbergs F.A.M.,
RA Carletti N., Sorg C., Odink K., Tarsay L., Wiesendanger W.,
RA de Wolf-Peters C., Shipcan R.;
RT "Restin: a novel intermediate filament-associated protein highly
  expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL ENBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92405160; PubMed=1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT "CLIP-170 links endocytic vesicles to microtubules.";
RL Cell 70:887-900(1992).
CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
  THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
  CYTOSKELETON.
CC -1- ALTERNATIVE PRODUCTS:
  Event=Alternative splicing; Named isoforms=2;
  Name=Long;
  IsoId=P30622-1; Sequence=Displayed;
  Name=Short;
  IsoId=P30622-2; Sequence=YSP_000765;
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
  OF HODGKIN'S DISEASE.
CC -1- SIMILARITY: Contains 2 CAP-Gly domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC
CC ENBL; X64838; CAA46050.1; -.
CC EMBL; M97501; AAA35693.1; -.
CC PIR; S22695; S22695.
CC Genew; HGNC:10461; RSN.
CC MIM; 179838; -.
CC GO; GO:0005768; C:Endosome; TAS.
CC GO; GO:0005882; C:Intermediate filament; TAS.
CC GO; GO:0015630; C:Microtubule cytoskeleton; TAS.
CC GO; GO:0008017; F:Microtubule binding; TAS.
CC GO; GO:0006899; P:nonselective vesicle transport; TAS.
CC InterPro; IPR000938; CAP-Gly.
```

```
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_GLY_2.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS00245; CAP_GLY_2; 2.
KW Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 78 120 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT DOMAIN 232 274 CAP-GLY 2.
FT DOMAIN 304 331 SER-RICH.
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
FT DOMAIN 1408 1421 CCHC-BOX.
FT VARSPIC 457 MISSING (in isoform Short).
FT /FTID=VSP_000765.
FT CONFLICT 1069 1069 D -> E (in REF. 2).
SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match 8.8%; Score 246; DB 1; Length 1427;
Best Local Similarity 24.4%; Pred. No. 0.0078;
Matches 114; Conservative 84; Mismatches 152; Indels 118; Gaps 20;

QY 133 KEPSKGDNTBEIROSDVGDR-----DHRPQEKKKKAGLQKE-----ITLQMQLN 180
DB 770 KASSEKSEMKKLROOLEAARKQIKHLEIKNAESKASSITRELOQREKLITNLQENLS 829
QY 181 TLS-----TPEKLAALCKKYAELLSE-----HNSOKMKLLQKKOSQL----- 220
DB 830 EVSQVKETLEKLOILKEKFAEASAEAVSVQVSQGETVVKLHKEQEFPNMLSSDLKLE 889
QY 221 -----VQEKDHLRGHSHKAVLARSKLESICRELQ-----HNRS-----LKRG 259
DB 890 NLADMEAKFREKDE---REEQLIKAKEKLENDIAEIMKMGDSNLSQITMNDLRLKRD 946
QY 260 VQAREEEERKREVTSHFQVTLNDIQLQWQ-HNERNSKLRQENNELAERLKLIDQVEL 318
DB 947 VEELQLKLTKANENASFLQKSTEDMTVKAROSQQAANKHBEKKELEKRLGSLDEKKMET 1006
QY 319 REEHIDKV-----FKHKO-----LQQLVDA--KQQAQE-----MLKEABE-RH 355
DB 1007 SHNQCELKARYRATSETKTHBEILQNLQTLDTEDKLKGASENSGLIQLEELRK 1066
QY 356 QREKDFLKEAVESQRMELMKQ-----ETHLKOQLALYTEKEPEFQNT--L 401
DB 1067 QADKAKAAQTABDAMQIMQIMQIMQIMQIMQIMQIMQIMQIMQIMQIMQIMQIMQIM 1126
QY 402 SKSESVFTTFQEMKMTKKIKCLEKETT-----MYRSRWESSNKA 442
DB 1127 NKSKELLTVENQKMEFPRKEITLQKAAQKSQLSALQENVKLABELGRSDEVTSHQ 1186
QY 443 LLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLQKRVQDLA 490
DB 1187 KLE--EERSVLNNQL--LEMK-KRESKFIKDADEEKASLOKSISITSA 1229

Search completed: June 8, 2004, 16:33:25
Job time : 20 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:27:31 ; Search time 45 Seconds
(without alignments)
3828.288 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 2785
Sequence: 1 MNQDKKNGAARQSNPKSSP.....APSTASGGTGPQRTSARA 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvrius:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2785	100.0	546	Q86YW3	Q86YW3 homo sapien
2	2780	99.8	546	Q86Y54	Q86Y54 homo sapien
3	2777	99.7	546	Q86Y86	Q86Y86 homo sapien
4	2769	99.4	546	Q86Y85	Q86Y85 homo sapien
5	1294	46.5	524	Q8BHN1	Q8BHN1 mus musculu
6	1284.5	46.1	528	Q8NUQ3	Q8NUQ3 homo sapien
7	1276	45.8	718	Q86Y52	Q86Y52 homo sapien
8	1275	45.8	715	Q8N313	Q8N313 homo sapien
9	1244	44.7	676	Q91969	Q91969 gallus gall
10	1229.5	44.1	684	Q8VBT1	Q8VBT1 mus musculu
11	1192	42.8	463	Q8BUK2	Q8BUK2 mus musculu
12	1008	36.2	505	Q8N382	Q8N382 homo sapien
13	646	23.2	241	Q8BPI1	Q8BPI1 mus musculu
14	598	21.5	116	Q86Y86	Q86Y86 homo sapien
15	597.5	21.5	515	Q9VBL3	Q9VBL3 drosophila
16	503	18.1	264	Q86LQ1	Q86LQ1 branchiosto

17	462.5	16.6	436	10	Q84VE3	Q84VE3 oryza sativ
18	457	16.4	241	11	Q8BWJ7	Q8BWJ7 mus musculu
19	451.5	16.2	404	10	Q8RXD7	Q8RXD7 arabidopsis
20	442.5	15.9	186	4	Q9P0X1	Q9P0X1 homo sapien
21	400.5	14.4	150	11	Q8C5K1	Q8C5K1 mus musculu
22	372.5	13.4	335	5	Q22666	Q22666 caenorhabdi
23	369.5	13.3	181	4	Q8BZA4	Q8BZA4 homo sapien
24	336	12.1	372	10	Q9FGD8	Q9FGD8 arabidopsis
25	291	10.4	1168	5	Q9VYU0	Q9VYU0 drosophila
26	281	10.1	900	3	Q9P3P5	Q9P3P5 neurospora
27	281	10.1	992	4	Q9NTH6	Q9NTH6 homo sapien
28	278.5	10.0	1455	4	Q9UPV0	Q9UPV0 homo sapien
29	261	9.4	2033	10	Q7XEH4	Q7XEH4 oryza sativ
30	260	9.3	1937	6	Q9TV62	Q9TV62 sus scrofa
31	258	9.3	1206	11	Q08815	Q08815 rattus norv
32	258	9.3	1941	13	Q9DGM4	Q9DGM4 gallus gall
33	257.5	9.2	610	11	Q9CS72	Q9CS72 mus musculu
34	257	9.2	1003	12	Q91LX9	Q91LX9 kaposi's sa
35	257	9.2	1944	13	Q9DGM5	Q9DGM5 gallus gall
36	255	9.2	1943	13	Q8JG72	Q8JG72 gallus gall
37	254.5	9.1	609	17	Q8TXA4	Q8TXA4 methanopyru
38	254.5	9.1	1089	12	Q40947	Q40947 kaposi's sa
39	254	9.1	1939	13	Q9PTY2	Q9PTY2 gallus gall
40	251.5	9.0	728	4	Q60561	Q60561 homo sapien
41	251.5	9.0	1005	4	Q9P216	Q9P216 homo sapien
42	251	9.0	1940	13	Q8AY28	Q8AY28 gallus gall
43	250.5	9.0	1233	11	Q54988	Q54988 mus musculu
44	250.5	9.0	2139	5	Q07569	Q07569 entamoeba h
45	250	9.0	1001	11	Q88664	Q88664 rattus norv

ALIGNMENTS

RESULT 1

ID Q86YW3 PRELIMINARY; PRT; 546 AA.
AC Q86YW3;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Taxilin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22448667; PubMed=12558796;
RA Nigami S., Satoh S., Nakano M., Shimizu H., Fukushima H., Maruyama A.,
RA Terano A., Shirataki H.;
RT "Taxilin; a novel syntaxin-binding protein that is involved in Ca2+-
RT dependent exocytosis in neuroendocrine cells.";
RL Genes Cells 8:17-28(2003).
DR ENBL; AF516206; AAC42465.1; --
SQ SEQUENCE 546 AA; 61891 MW; 698CD74F78897DF6 CRC64;

Query Match 100.0%; Score 2785; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.1e-124;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKNQDKKNGAARQSNPKSSPQPEAGPEGAOERPSQAAPVAEAGPGSSQAEPKPKPGAQA	60
Db	1	MKNQDKKNGAARQSNPKSSPQPEAGPEGAOERPSQAAPVAEAGPGSSQAEPKPKPGAQA	60
Qy	61	RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPDAEKSRITYARN	120
Db	61	RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPDAEKSRITYARN	120
Qy	121	GPPEPTPVVNGSKERPSKGPDPNTEINQSGVGRDHRHPQEKKAIGLKGKETILLMOTLN	180
Db	121	GPPEPTPVVNGSKERPSKGPDPNTEINQSGVGRDHRHPQEKKAIGLKGKETILLMOTLN	180


```

181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
241 LESLCRELQRNRSLSKEGVQARAREBEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
241 LESLCRELQRNRSLSKEGVQARAREBEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
301 ENMELAEERLKKLIQOYELREBEHDKVFKHKLQOQLVDKLAQQAQEMLKAEERHOREKD 360
301 ENMELAEERLKKLIQOYELREBEHDKVFKHKLQOQLVDKLAQQAQEMLKAEERHOREKD 360
361 FLKKEAVESQRMCELMKQOETHLKKQALALYATEKFEFQNTLSKSEVFITFKQEMEKMTK 420
361 FLKKEAVESQRMCELMKQOETHLKKQALALYATEKFEFQNTLSKSEVFITFKQEMEKMTK 420
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCYPGAPSTASQGTGPQE 540
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCYPGAPSTASQGTGPQE 540
541 PTSARA 546
541 PTSARA 546

RESULT 2
Q86T54 PRELIMINARY; PRT; 546 AA.
ID Q86T54
AC Q86T54;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE Hypothetical protein DKEZP451K215.
GN DKEZP451K215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=human skeletal muscle;
RA Koerner K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832338; CAD91138.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728D674F78897DF6 CRC64;

Query Match 99.8%; Score 2780; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.8e-124;
Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKEGAQA 60
DB 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKEGAQA 60
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
DB 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
QY 121 GEPEPTPVNNGEKPSPKGDPTNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
DB 121 GEPEPTPVNNGEKPSPKGDPTNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
QY 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
DB 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240

```

```

241 LESLCRELQRNRSLSKEGVQARAREBEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
241 LESLCRELQRNRSLSKEGVQARAREBEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
301 ENMELAEERLKKLIQOYELREBEHDKVFKHKLQOQLVDKLAQQAQEMLKAEERHOREKD 360
301 ENMELAEERLKKLIQOYELREBEHDKVFKHKLQOQLVDKLAQQAQEMLKAEERHOREKD 360
361 FLKKEAVESQRMCELMKQOETHLKKQALALYATEKFEFQNTLSKSEVFITFKQEMEKMTK 420
361 FLKKEAVESQRMCELMKQOETHLKKQALALYATEKFEFQNTLSKSEVFITFKQEMEKMTK 420
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCYPGAPSTASQGTGPQE 540
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCYPGAPSTASQGTGPQE 540
541 PTSARA 546
541 PTSARA 546

RESULT 3
Q86T86 PRELIMINARY; PRT; 546 AA.
ID Q86T86
AC Q86T86;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE Hypothetical protein DKEZP451I0918.
GN DKEZP451I0918.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ausorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832636; CAD89951.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; FEBCAD4F753F7FF1 CRC64;

Query Match 99.7%; Score 2777; DB 4; Length 546;
Best Local Similarity 99.6%; Pred. No. 2.5e-124;
Matches 544; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKEGAQA 60
DB 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKEGAQA 60
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
DB 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
QY 121 GEPEPTPVNNGEKPSPKGDPTNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
DB 121 GEPEPTPVNNGEKPSPKGDPTNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
QY 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
DB 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMKLLQKQSQLVQVNDIQLQMEQHNRNSKLQ 240
241 LESLCRELQRNRSLSKEGVQARAREBEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
241 LESLCRELQRNRSLSKEGVQARAREBEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
301 ENMELAEERLKKLIQOYELREBEHDKVFKHKLQOQLVDKLAQQAQEMLKAEERHOREKD 360

```



```

Db 301 ENNELAERLKKLIEQVELREEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQREKD 360
Qy 361 FLKKEAVSORMCMLMKQOETHLKQALALYTKFEFFQNTLSKSSSEVFTTFKQEMEKMTK 420
Db 361 FLKKEAVSORMCMLMKQOETHLKQALALYTKFEFFQNTLSKSSSEVFTTFKQEMEKMTK 420
Qy 421 KIKKLEKETTMYRSRWESSNKALLEAEKTVRDKEGLEGLQVKIQLEKLCRALQTERND 480
Db 421 KIKKLEKETTMYRSRWESSNKALLEAEKTVRDKEGLEGLQVKIQLEKLCRALQTERND 480
Qy 481 LNKRVODLSAGOGSLTDSGPRRPGCAQAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Db 481 LNKRVODLSAGOGSLTDSGPRRPGCAQAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Qy 541 PTSARA 546
Db 541 PTSARA 546

RESULT 4
Q86T85 PRELIMINARY; PRT; 546 AA.
AC Q86T85;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE Hypothetical protein DK92p45J0118.
GN DK92P45J0118.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansoerge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL32637; CAD89952.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61846 MW; 36718BAE3A7B6C2 CRC64;

```

Query Match 99.4%; Score 2769; DB 4; Length 546;
 Best Local Similarity 99.5%; Pred. No. 6.1e-124;
 Matches 543; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 MNQDKKNGAANKSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPEGAQA 60
Db 1 MNQDKKNGAANKSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPEGAQA 60
Qy 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNOGGPGEDGAGQGEPAEPDAEKSRITYVARN 120
Db 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNOGGPGEDGAGQGEPAEPDAEKSRITYVARN 120
Qy 121 GPEPTPVVNGKEPKSGDPNTEIRQSEVGDGRDHRPQEKKAAGLKGITLLMOTLN 180
Db 121 GPEPTPVVNGKEPKSGDPNTEIRQSEVGDGRDHRPQEKKAAGLKGITLLMOTLN 180
Qy 181 TLSTPEKLAALCKKYAEELLEHRNSQKMLQKQOSQVQKDHLEHSEKAVLARSK 240
Db 181 TLSTPEKLAALCKKYAEELLEHRNSQKMLQKQOSQVQKDHLEHSEKAVLARSK 240
Qy 241 LESLCELQRHNRSLKEEGVQARAREEKKRKEVTSHPQVTLNDIQLOMEQHNRNSKLQR 300
Db 241 LESLCELQRHNRSLKEEGVQARAREEKKRKEVTSHPQVTLNDIQLOMEQHNRNSKLQR 300
Qy 301 ENNELAERLKKLIEQVELREEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQREKD 360
Db 301 ENNELAERLKKLIEQVELREEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQREKD 360
Qy 361 FLKKEAVSORMCMLMKQOETHLKQALALYTKFEFFQNTLSKSSSEVFTTFKQEMEKMTK 420
Db 361 FLKKEAVSORMCMLMKQOETHLKQALALYTKFEFFQNTLSKSSSEVFTTFKQEMEKMTK 420

```

```

Qy 421 KIKKLEKETTMYRSRWESSNKALLEAEKTVRDKEGLEGLQVKIQLEKLCRALQTERND 480
Db 421 KIKKLEKETTMYRSRWESSNKALLEAEKTVRDKEGLEGLQVKIQLEKLCRALQTERND 480
Qy 481 LNKRVODLSAGOGSLTDSGPRRPGCAQAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Db 481 LNKRVODLSAGOGSLTDSGPRRPGCAQAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Qy 541 PTSARA 546
Db 541 PTSARA 546

RESULT 5
Q86HN1 PRELIMINARY; PRT; 524 AA.
AC Q86HN1;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN RBBP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030100; BAC26785.1; -
DR EMBL; AK031783; BAC27547.1; -
DR EMBL; AK044130; BAC31791.1; -
DR MGD; MGI:1194910; Rbbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016564; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P. .; IDA.
KW Hypothetical protein.
SQ SEQUENCE 524 AA; 60308 MW; 022877633B4ED7C CRC64;

```

Query Match 46.5%; Score 1294; DB 11; Length 524;
 Best Local Similarity 53.9%; Pred. No. 5.2e-54;
 Matches 265; Conservative 80; Mismatches 89; Indels 58; Gaps 5;

```

Qy 7 KNGAAGKSNPKSPGQPEAGPE---GAQERPSQAAPAVEAEGPGSSQAP---RKPEGAQ 59
Db 12 RGGTTEASEGGRGRRRSPQKTEIGTWEEARICGLGVKADWVCNSQANDILQHDPSQ 71
Qy 60 ARTAQSGALR--DVSEELSRQLEDILSTYCV-DNNQGGPGEDGAGQGEPAEPDAEKSRITY 116
Db 72 GGTTKHSLGDEGSDFTTKRNLVSSVFTQEKREETPGREARTGPPDGGQDSECSR-- 129
Qy 117 VARNGEPEPTPVVNGKEPKSGDPNTEIRQSEVGDGRDHRPQEKKAAGLKGITLLM 176
Db 130 -----NKEKTLKEVLLM 143
Qy 177 QTLNTLSTPEKLAALCKKYAEELLEHRNSQKMLQKQOSQVQKDHLEHSEKAVL 236
Db 144 QALNTLSTPEKLAALCKKYADLLEESNRNVQKMLQKQOIVKKEVHLQSEHSAK 203
Qy 237 ARSKLESICRLQRHNRSLKEEGVQARAREEKKRKEVTSHPQVTLNDIQLOMEQHNRNS 296
Db 204 ARSKLESICRLQRHNRSLKEEGVQARAREEKKRKEVTSHPQVTLNDIQLOMEQHNRNS 263
Qy 297 KLRQENMELAEKLCLEQVELREEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQ 356
Db 264 KLRQENIELGSKLCKLIEQVALREEHIDKVFHKDLQOQLVDAKLOQTTLIKEADERHQ 323

```

```

QY 357 REKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFQNTLSKSESVFTTFQEMBE 416
DQ 324 REREFFLLKEATSRHRYEQOMQOEVQLKQOOLSMDYDFEFQNTMAKSNELFPTTFQEMBE 383
QY 417 KMTKKIKKLEKSTMTYRSRSESSNKALLEMAEAKTVDRDLEGLQVKIQRLEKLCRALQOT 476
DQ 384 KMTKKIKKLEKSTMTYRSRSESSNKALLEMAEAKTVDRDLEGLQVKIQRLEKLCRALQOT 443
QY 477 ERNDLANKRVQDL 488
DQ 444 ERNELNEKVEVL 455

RESULT 6
Q9NUQ3 PRELIMINARY; PRT; 528 AA.
ID Q9NUQ3
AC Q9NUQ3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical protein FLJ11209.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isagai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wegasuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT *NEDO human cDNA sequencing project.*;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EXBL; AK002071; BAA92068.1;
DR Genew; HGNC:18578; CXorf15.
KW Hypothetical protein.
SQ SEQUENCE 528 AA; 60605 MW; CAE8D781DE06ACB3 CRC64;

```

```

Query Match 46.1%; Score 1284.5; DB 4; Length 528;
Best Local Similarity 53.4%; Pred. No. 1.5e-53;
Matches 285; Conservative 75; Mismatches 131; Indels 43; Gaps 10;

QY 28 EGAQERPSQAAPVAEAGPGSSQAPRK--PEGAQARTAGSGLRDVSELSRQLEDILT 85
DQ 7 EAARGGGGAEEATEAGRGRRRSRQKFEIGTWERAGICGLGVKADMLCNSQSDILQH 66
QY 86 YCVDRNQGPGEGAQGEPAEPDAEKSRITYVARNGEPEPTPVNKEKPSKGDPTNTEI 145
DQ 67 Q--GSNCGGTSNKSLEEDSGSDFITENRNLS-----PAYCTQSRBEIPGG---BA 114
QY 146 RQSDVGDHRHRRPQEKKAAGKGGKITTILMQTLTSTPEEKLAALCKKYAELEEHEN 205
DQ 115 RTDPPDQDSECNKKEKT--LGKEVLLMQALNTLSTPEEKLAALCKKYADLLEERS 172
QY 206 SQQMKLQKQOQLVOEKHLEHSHKAVLARSKLESLCRELQNRNRSKMBGVORARE 265
DQ 173 VQKQKILQKQOQVKEKVKHLQSEHSKAILARSKLESLCRELQNRNRSKMBGVORARE 232
QY 266 EEBKREKVTSHFQVTLNDILOMEQHNRNSKLQENMELAEKILKQIYELREEHIDK 325
DQ 233 EEBRKEATNHFQITLDEIQALQEQHDIENAKLRQENIELGEVKKQIYELREEHIDK 292
QY 326 VFKHKLQOOLVDAKLOQAQEMKKEAEERHQRKDFLLKEAVESQRMCELMKQOETHLQ 385
DQ 293 VFKRKLQOOLVDAKLOQTTLTQTLKEADEKHQREREFLLKEATSRHRYEQOMQOEVQLK 352
QY 386 QLALYTEKFEFQNTLSKSESVFTTFQEMBEKMTKKIKKLEKTTMYRSRSESSNKALLE 445
DQ 353 QLSLYMDKFEFQNTMAKSNELFPTTFQEMBEKMTKKIKKLEKSTMTYRSRSESSNKALLE 412

```

```

QY 446 MABEKTVDKXLEGLQVKIQRLEKLCRALQOTERNDLNKRQV-----DL 488
DQ 413 MABEKTVDKXLEGLQVKIQRLEKLCRALQOTERNELNEKVEVLKQVSIKAAIKANRDL 472
QY 489 SAGGGSGLT--DSGPE-----RRPBGGAQA--PSSPR--VTEAPCYPGAPSTEA 532
DQ 473 ATPVMQPTALDSHKELTNTSSKRGALGAHLEAPKQSRSAVQKQPPSTGSAPIES 526

RESULT 7
Q86TS2 PRELIMINARY; PRT; 718 AA.
ID Q86TS2
AC Q86TS2
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKFZp451G083 (Fragment).
GN DKFZp451G083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wisemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EXBL; AL831988; CAD91140.1;
KW Hypothetical protein.
SQ SEQUENCE 718 AA; 80290 MW; 16C3BF22C19559A4 CRC64;

```

```

Query Match 45.8%; Score 1276; DB 4; Length 718;
Best Local Similarity 45.9%; Pred. No. 5.1e-53;
Matches 277; Conservative 94; Mismatches 141; Indels 92; Gaps 10;

QY 3 NODKNGAAGKSNPKSPGQPEAGP--EGAQERPSQAAPVAEAGPGSSQAPRKPEGQA 60
DQ 38 NHSEQLSAEROSTP---PQDSSSLPSHNGLEKEDQ-----DSFTPVQPEK----- 81
QY 61 RTAQSGALRDVSELSRQLEDILTSTYCVDRNQGPGEGAQGEPAEPDAEKSRITYVARN 120
DQ 82 ---EASVDPDISELNRQLEDILNTY--GSAASTAGKGSARASEQPNASPDN--ED 133
QY 121 GEPEPTPVNNGEKEPSKGDPTNTEIYQSDVGDHRHRRPQEKKAAGKGGKITTILMQTLN 180
DQ 134 GDCBETTEAGREPVASGEPTVK---EPVSNKEQK--LEKCKILKGLGKEANLLMQNLN 187
QY 181 TLSTPEEKLAALCKKYAELEEHENRSQKMKLQKQOQLVOEKHLEHSHKAVLARSK 240
DQ 188 KLQTPPEKFDLFKKYASLLDEHRTQKKLQKQOQVIOKEKQQLQGEHSHRALLARSK 247
QY 241 LESLCRELQNRNRSKMBGVORAREEEKREKVTSHFQVTLNDILOMEQHNRNSKLQ 300
DQ 248 LESLCRELQNRNRSKMBGVORAREEEKREKVTSHFQVTLNDILOMEQHNRNSKLQ 307
QY 301 ENMELAEKILKQIYELREEHIDKVPFKHKLQOOLVDAKLOQAQEMKKEAEERHQRK 360
DQ 308 ENTELAELKSIIDQYELREEHIDKVPFKHKLQOOLVDAKLOQAQEMKKEAEERHQRK 367
QY 361 FLKKEAVESQRMCELMKQOETHLKOALYTEKFEFQNTLSKSESVFTTFQEMBEKMTK 420
DQ 368 YLLNQAAEWKILQAKMLKEQETVQLQALTYLSGRFEFQSTLTSTKSNVETATFQEMDKTK 427
QY 421 KIKKLEKTTMYRSRSESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQOTERN 480
DQ 428 NMKLEKQDTATWKARFENCNALLDAMIIEKALRAKVECFVWKIGRLENLCRALQOERNE 487
QY 481 LNKRVQDLSAGGGSGLTDSGPERPEG----- 507
DQ 488 LHKIRDAEISEKDDQSQFNSDERPESNVSDQEDDAEYNSVQTAVQNQLATAFMIIEHP 547

```

```

QY 508 -----PGQAQSPSP-----RVTEAPCYPCAPSTEASGOTGPOE 540
Db 548 ESTPHOSKETQPEIGSSQESADAALKEPEQPLIPSRDSESLPPLTPOAEAGGSDAEP 607
QY 541 PTSA 544
Db 608 PSKA 611

RESULT 8
Q8N3L3
ID Q8N3L3 PRELIMINARY; PRT; 715 AA.
AC Q8N3L3;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein DKF2p451f022 (fragment).
GN DKF2p451f022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834248; CAD38924.2; -
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 715 AA; 79798 MW; 11C50191BCD26582 CRC64;

Query Match 45.8%; Score 1275; DB 4; Length 715;
Best Local Similarity 45.9%; Pred. No. 5.7e-53;
Matches 277; Conservative 94; Mismatches 141; Indels 92; Gaps 10;

QY 3 NODKNGAALKQSNPKSPGQPEAGP--EQAQERPSQAAPAVEAGPGSSQAQPKPEGAQA 60
Db 35 NHBSEQLSAERQSTP---PGDSSSLPSHNGLEKEDGQ-----DSPTVPQPEK----- 78
QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGPAPPEAERKSRVTVARN 120
Db 79 ---EASVHPDISELNRLQLEDIINTY--GSAASTAGKEGSRASQEPENAESPDN---ED 130
QY 121 GEPEPTPVNGEKEPSKGDPTNEIRQSDVGDGRDHRRRPOEKKKAGLGLKEITLLMOTLN 180
Db 131 GDCEETTEAGRPVAGSEPTVK---EPVSNKEQK--LEKKILKGLGKEANLLMQLN 184
QY 181 TLSTPEKLAALCKKYAELLERHNSQKQMLLOKQSQOLVQEKDHLRGEHSHKAVLARSK 240
Db 185 KLQTPREKDFLPKCYAELLERHNSQKQMLLOKQSQOLVQEKDHLRGEHSHKAVLARSK 244
QY 241 LESLCELOHNSLSKEEGVQARAREEERKEVTSHPQVTLNDIOLQMEQHNSKLNKLRQ 300
Db 245 LESLCELOHNSKTLKEALQAREEERKEVTSHPQVTLNDIOLQMEQHNSKLNKLRQ 304
QY 301 ENNELAERLKLIEQVELREEHIDKVPFKHNDLQQLVDAKLOQAQEMLEAEERHOREKD 360
Db 305 ENTELAEKLSIIOVELREEHLDKIFKHREELQQLVDAKLEQSQEMKAEERHOREKKE 364
QY 361 FLKEAVESQRCMELKQOETHLKOQALYTEKPEFQVTLSSSEVFTTFKQEMKMTK 420
Db 365 YLLNQAEAKLQAKVLEQSTVLOAQVTLSSSEVFTTFKQEMKMTK 424
QY 421 KIKKLEKTTWRSWESSNKALLEMAEETKVRDKELEGVQVIRLEKLCRALQTERND 480
Db 425 KMKKLEKDTATWARKFENCNKALLDMLLEKALRAKVECFVWKIGLENLCRALQBERNE 484
QY 481 INTRVQDLSAGGQGSITSDSPERRPEG----- 507
Db 485 LHKKIRDAEISSEKDDQSQHNSDEEFPNSVSDVIDAEVNSVQTAVKNLATAFMIHHP 544

```

```

QY 508 -----PGQAQSPSP-----RVTEAPCYPCAPSTEASGOTGPOE 540
Db 545 ESTPHOSKETQPEIGSSQESADAALKEPEQPLIPSRDSESLPPLTPOAEAGGSDAEP 604
QY 541 PTSA 544
Db 605 PSKA 608

RESULT 9
Q9I969
ID Q9I969 PRELIMINARY; PRT; 676 AA.
AC Q9I969;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DE 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle derived protein.
GN MDP77.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Muscle;
RX MEDLINE=20175243; PubMed=10708594;
RA Uyeda A., Fukui I., Fujimori K., Kiyosue K., Nishimune H., Kasai M.,
RA Taguchi T.;
RT "MDP77: A novel neurite-outgrowth-promoting protein predominantly
RT expressed in chick muscles.";
RL Biochem. Biophys. Res. Commun. 269:564-569(2000).
DR EMBL; D89999; BAA94755.1; -
DR PIR; JC7222; JC7222.
SQ SEQUENCE 676 AA; 77020 MW; FCEA9E393250EE94 CRC64;

Query Match 44.7%; Score 1244; DB 13; Length 676;
Best Local Similarity 54.3%; Pred. No. 1.6e-51;
Matches 260; Conservative 81; Mismatches 104; Indels 34; Gaps 8;

QY 12 KQSNPKSPGQPEAGPQERPSQAAPAVEAGPGSSQAQPKPEGAQAQTAQSGALRDV 71
Db 9 KQCVTTFTQDNQG-----QSKAEVFSQPLSP-TNQTSAQPEWATC-----DI 53
QY 72 SEELSRQLEDILSTYCVNNQGGPGEDGA-----QGEPAPEPAERKSRVTVARNGEPPTP 127
Db 54 SEELNRQLEDIITKY-----GSAASLVEKEGTTATDTPKEDVGSMDAECED-- 102
QY 128 VYNGEKEPSKGDPTNEIRQSDVGDGRDHRRRPOEKKKAGLGLKEITLLMOTLNTLSTPEE 187
Db 103 -VNESEHDKKPAFG-DASRAKPSASKQK--LEKKILKGLGKEATLLMQSLNKLTPEE 158
QY 188 KLAALCKKYAELLERHNSQKQMLLOKQSQOLVQEKDHLRGEHSHKAVLARSKLSLCRE 247
Db 159 KLDLLPKCYAELLERHNSQKQMLLOKQSQOLVQEKDHLRGEHSHKAVLARSKLSLCRE 218
QY 248 IQRHNSLSKEEGVQARAREEERKEVTSHPQVTLNDIOLQMEQHNSKLNKLNELAE 307
Db 219 LQHNKTLKEBTQARAREEERKEVTSHPQVTLNDIOLQMEQHNSKLNKLNELAE 278
QY 308 RLKKLIEQVELREEHIDKVPFKHNDLQQLVDAKLOQAQEMLEAEERHOREKDFLKEAV 367
Db 279 KLSIIOVELREEHLDKIFKHREELQQLVDAKLEQSQEMKAEERHOREKDFLKEAV 338
QY 368 ESQRCMELKQOETHLKOQALYTEKPEFQVTLSSSEVFTTFKQEMKMTKIKKLEK 427
Db 339 EWLQAKMLKEQSTVLOAQVTLSSSEVFTTFKQEMKMTKIKKLEK 398
QY 428 ETTWRSWESSNKALLEMAEETKVRDKELEGVQVIRLEKLCRALQTERNDLNRVQ 486
Db 399 DFTATWRSWESSNKALLEMAEETKVRDKELEGVQVIRLEKLCRALQTERNDLNRVQ 457

```

Search completed: June 8, 2004, 16:34:22
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:37:57 ; Search time 50 Seconds
(without alignments)
3072.218 Million cell updates/sec

Title: US-10-023-529-44
Perfect score: 546
Sequence: 1 MNQDKKNGAAQSNPKSSP.....APSTRASGTGQPEPTSA 546

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1155919 seqs, 281338677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	546	100.0	546	9	US-09-976-740-44
2	546	100.0	546	12	US-10-023-529-44
3	546	100.0	546	13	US-10-023-529-44
4	546	100.0	546	13	US-10-023-529-44
5	546	100.0	546	15	US-10-023-529-44
6	476	87.2	530	12	US-10-276-774-2134
7	429	78.6	530	9	US-09-962-055-8
8	429	78.6	530	9	US-09-976-740-8
9	429	78.6	530	12	US-10-023-529-8
10	429	78.6	530	13	US-10-023-529-8
11	429	78.6	530	13	US-10-023-529-8
12	429	78.6	530	15	US-10-023-529-8
13	172	31.5	557	9	US-09-962-055-5
14	172	31.5	557	9	US-09-976-740-5
15	172	31.5	557	12	US-10-023-529-5

16	172	31.5	557	13	US-10-023-529-5	Sequence 5, Appli
17	172	31.5	557	13	US-10-023-529-5	Sequence 5, Appli
18	172	31.5	557	15	US-10-023-529-5	Sequence 5, Appli
19	17	3.1	85	15	US-10-023-529-5	Sequence 3413, Ap
20	11	2.0	436	12	US-10-023-529-5	Sequence 207434,
21	9	1.6	99	9	US-09-764-869-1082	Sequence 1082, Ap
22	9	1.6	99	14	US-10-091-504-1082	Sequence 1082, Ap
23	9	1.6	99	15	US-10-237-577-1082	Sequence 112, App
24	9	1.6	594	9	US-09-801-368-112	Sequence 1762, Ap
25	9	1.6	594	15	US-10-369-493-1762	Sequence 72115, A
26	9	1.6	637	12	US-10-282-122A-72115	Sequence 130, App
27	8	1.5	25	16	US-10-443-622-130	Sequence 48446, A
28	8	1.5	106	12	US-10-425-114-48446	Sequence 272074,
29	8	1.5	121	12	US-10-424-599-272074	Sequence 230732,
30	8	1.5	148	12	US-10-424-599-230732	Sequence 122, App
31	8	1.5	156	15	US-10-443-622-122	Sequence 316, App
32	8	1.5	229	14	US-10-080-170-316	Sequence 12433, A
33	8	1.5	413	14	US-10-136-761-12433	Sequence 21212, A
34	8	1.5	447	15	US-10-369-493-21212	Sequence 7865, Ap
35	8	1.5	450	12	US-10-335-977-7865	Sequence 6430, Ap
36	8	1.5	451	12	US-10-335-977-7866	Sequence 23425, A
37	8	1.5	484	9	US-09-738-626-6430	Sequence 41, Appl
38	8	1.5	525	15	US-10-359-493-23425	Sequence 41, Appl
39	7	1.3	7	9	US-09-962-055-41	Sequence 41, Appl
40	7	1.3	7	9	US-09-976-740-41	Sequence 41, Appl
41	7	1.3	7	12	US-10-023-529-41	Sequence 41, Appl
42	7	1.3	7	13	US-10-023-529-41	Sequence 41, Appl
43	7	1.3	7	13	US-10-023-529-41	Sequence 41, Appl
44	7	1.3	7	15	US-10-616-187-41	Sequence 2071, Ap
45	7	1.3	18	14	US-10-225-567A-2071	

ALIGNMENTS

RESULT 1

US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44

Query Match 100.0%; Score 546; DB 9; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNQDKKNGAAQSNPKSSPQPEAGPQAPVAVGPGSSQAPKPEGAQA 60
DB 1 MNQDKKNGAAQSNPKSSPQPEAGPQAPVAVGPGSSQAPKPEGAQA 60